

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:49:21 ; Search time 69.1132 Seconds  
(without alignments)  
4217.989 Million cell updates/sec

Title: US-09-965-553-1  
Perfect score: 18  
Sequence: 1 tggcgcggttgcacatg 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rnd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	129	14	B0853899
2	18	100.0	166	17	AL771403
3	18	100.0	177	17	AL757122
4	18	100.0	178	17	AQ958189
5	18	100.0	199	17	BH243164
6	18	100.0	216	17	CNS00711

Result No.	Score	Query Match	Length	DB ID	Description
7	18	100.0	263	17	BH243803
8	18	100.0	388	17	B62585
9	18	100.0	401	17	BH244138
10	18	100.0	410	17	BH244101
11	18	100.0	415	17	BH243289
12	18	100.0	418	17	BH243156
13	18	100.0	419	17	BH252783
14	18	100.0	422	17	BH244105
15	18	100.0	424	17	BH244079
16	18	100.0	437	17	BH244067
17	18	100.0	439	17	BH243367
18	18	100.0	444	17	BH243576
19	18	100.0	446	17	BH235368
20	18	100.0	451	17	CNS00M07
21	18	100.0	452	17	CNS00M07
22	18	100.0	455	17	CNS00R39
23	18	100.0	456	14	BQ989005
24	18	100.0	457	17	AQ958592
25	18	100.0	460	17	B95970
26	18	100.0	470	17	BH235524
27	18	100.0	471	17	CNS00MSM
28	18	100.0	478	17	BH243935
29	18	100.0	483	17	CNS00MGM
30	18	100.0	486	17	B23757
31	18	100.0	493	17	BH243182
32	18	100.0	493	17	BH243582
33	18	100.0	493	17	BH243774
34	18	100.0	505	17	CNS00O0V
35	18	100.0	506	17	B27265
36	18	100.0	514	17	BH243704
37	18	100.0	520	17	CNS00R7E
38	18	100.0	521	17	CNS00R25
39	18	100.0	522	17	BH244760
40	18	100.0	524	17	BH243951
41	18	100.0	527	14	BQ851421
42	18	100.0	528	17	BH244809
43	18	100.0	537	17	BH235380
44	18	100.0	543	17	CNS00NS4
45	18	100.0	552	17	BH235544

## ALIGNMENTS

RESULT 1  
B0853899/c  
LOCUS  
DEFINITION  
OCB21M16, yg\_ab1 OG\_ABCDI lettuce salinas Lactuca sativa cDNA clone  
B0853899  
ACCESSION  
B0853899.1 GI:22239364  
VERSION  
B0853899.1  
KEYWORDS  
SOURCE  
ORGANISM  
EST  
Lactuca sativa  
Lactuca sativa  
Lactuca sativa  
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Asteridae: euasterids II; Asterales: Asteraceae: Lactuceae: Lactuca.  
1 (bases 1 to 129)  
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L., and Bradford, K.  
Lettuce and Sunflower ESTs from the Compositae Genome Project  
http://compgenomics.ucdavis.edu/  
Unpublished (2002)  
Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]

belongs to contig OG\_CA.Contig5556, see <http://cgpdb.ucdavis.edu/> for details.  
 Plate: OG821 row: M column: 16.  
 Location/Qualifiers

# FEATURES

source

1.129  
 /organism="Lactuca sativa"  
 /cultivar="Salinas"  
 /db\_xref="taxon:4236"  
 /clone="OG821M16"  
 /clone\_1lb="OG-ABCDI lettuce salinas"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNA5f1AB: The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformationally made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/TAG-LIB-OG-ABCDI-lettuce-salinas>  
 TAG-TISSUE-chemical induction  
 TAG-SBO-TGTRGCCGGC

# BASE COUNT

41 a 35 c 22 g 31 t

# ORIGIN

Query Match 100.0%; Score 18; DB 14; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGCGCCGCTTGCCCAATTG 18  
 ||||||||||||||||  
 Db 76 TGCGCCGCTTGCCCAATTG 59

# RESULT 2

AL771403

LOCUS AL771403 166 bp DNA linear GSS 19-JUN-2002  
 DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-179H03-013563,  
 genomic survey sequence.

# ACCESSION

AL771403.1 GI:21533605

# VERSION

AL771403.1 GI:21533605

# KEYWORDS

GSS.

# SOURCE

thale cress.

# ORGANISM

Arabidopsis thaliana

# REFERENCE

1 Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Saedler,H.

# AUTHORS

and Weissshaar,B.

# TITLE

A pipeline for automated high-throughput generation of FSTs

# JOURNAL

Unpublished

# REFERENCE

2 Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weissshaar,B.

# AUTHORS

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)

# TITLE

for flanking sequence tag based reverse genetics

# JOURNAL

Unpublished

# REFERENCE

3 (bases 1 to 166)

# AUTHORS

Li,Y., Rosso,M., Strizhov,N. and Weissshaar,B.

# JOURNAL

Submitted (17-JUN-2002) Weissshaar B., Max-Planck-Institut fuer

# COMMENT

This sequence is recovered from the left border of the T-DNA. It

# FEATURES

indicates an insertion within the locus defined by clone F2316. The

# FEATURES

sequences are generated at the MPI for Plant Breeding Research in

# FEATURES

the context of the GABI-Kat project. GABI-Kat is part of the German

# FEATURES

Plant Genomics program designated 'GABI'. Information on line

# FEATURES

availability can be found at:

# FEATURES

<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

source

1.166  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone="GK-179H03-013563"  
 /clone\_1lb="Arabidopsis thaliana T-DNA insertion lines"  
 /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAB161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

# BASE COUNT

49 a 30 c 34 g 53 t

# ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGCGCCGCTTGCCCAATTG 18  
 ||||||||||||||||  
 Db 145 TGCGCCGCTTGCCCAATTG 162

# RESULT 3

AL757122/c

LOCUS AL757122 177 bp DNA linear GSS 18-JUN-2002  
 DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-118B03-012518,  
 genomic survey sequence.

# ACCESSION

AL757122

# VERSION

AL757122.1 GI:21495470

# KEYWORDS

GSS.

# SOURCE

thale cress.

# ORGANISM

Arabidopsis thaliana

# REFERENCE

1 Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Saedler,H.

# AUTHORS

and Weissshaar,B.

# TITLE

A pipeline for automated high-throughput generation of FSTs

# JOURNAL

Unpublished

# REFERENCE

2 Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weissshaar,B.

# AUTHORS

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)

# TITLE

for flanking sequence tag based reverse genetics

# JOURNAL

Unpublished

# REFERENCE

3 (bases 1 to 177)

# AUTHORS

Li,Y., Rosso,M., Strizhov,N. and Weissshaar,B.

# JOURNAL

Submitted (17-JUN-2002) Weissshaar B., Max-Planck-Institut fuer

# COMMENT

This sequence is recovered from the left border of the T-DNA. It

# FEATURES

indicates an insertion close to or within gene At1g38440. The

# FEATURES

sequences are generated at the MPI for Plant Breeding Research in

# FEATURES

the context of the GABI-Kat project. GABI-Kat is part of the German

# FEATURES

Plant Genomics program designated 'GABI'. Information on line

# FEATURES

availability can be found at:

# FEATURES

<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

# FEATURES

Location/Qualifiers

# FEATURES

1.177

# FEATURES

/organism="Arabidopsis thaliana"

# FEATURES

/strain="Columbia 0"

# FEATURES

/db\_xref="taxon:3702"

# FEATURES

/clone="GK-118B03-012518"

# FEATURES

/clone\_1lb="Arabidopsis thaliana T-DNA insertion lines"

# FEATURES

/note="PCR was performed on DNA from Arabidopsis thaliana

# FEATURES

plants (T1) which were transformed with the T-DNA from



vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the *A. thaliana* nuclear genome sequence were processed for submission. T-DNA derived sequences were removed.

BASE COUNT 54 a 29 c 42 g 52 t  
ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 177;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCGCTGCCAATTG 18  
|||||  
DB 34 TGGCGCGCTGCCAATTG 17

RESULT 4  
LOCUS A0958189 178 bp DNA linear GSS 28-JAN-2000  
DEFINITION LERAV83TF LERA Arabidopsis thaliana genomic clone LERAV83, DNA sequence.

ACCESSION A0958189  
VERSION A0958189  
KEYWORDS GSS.  
SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 178)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

AUTHORS 1 (bases 1 to 178)  
Buell, C.R., Lin, X., Pal, G., Barnstead, M., Bowman, C., Uterbach, T., Feldblyum, T., Liang, F., Greasy, T., and Fraser, C.M.

TITLE Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms

JOURNAL Unpublished (2000)

COMMENT Contact: Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: atetigr.org

For additional information, see <http://www.tigr.org/tdb/at/at.html>  
Seq primer: TF  
Class: shotgun.

FEATURES  
source Location/Qualifiers  
1..178

/organism="Arabidopsis thaliana"  
/strain="Landsberg erecta"  
/db\_xref="taxon:3702"

/clone="LERAV83"

/clone.lib="LERA"

/note="Organ: leaf; Vector: PHOS1; Total genomic DNA was sheared to 0.9-1 kbp before ligation."

BASE COUNT 49 a 26 c 30 g 73 t  
ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 178;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCGCTGCCAATTG 18  
|||||  
DB 77 TGGCGCGCTGCCAATTG 94

RESULT 5  
LOCUS BH243164 199 bp DNA linear GSS 13-NOV-2001  
DEFINITION AUHHC32TR AUH Arabidopsis thaliana genomic clone AUHHC32, DNA

sequence.

ACCESSION BH243164  
VERSION BH243164.1 GI:16918951  
KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 199)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

AUTHORS 1 (bases 1 to 199)  
Town, C.D., Whitelaw, C.A., Pal, G., Van Aken, S.E., Uterbach, T.V., Feldblyum, T.V., and Fraser, C.M.

TITLE Survey sequencing of Arabidopsis thaliana BAC F1307

JOURNAL Unpublished (2001)

COMMENT Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtownetigr.org  
From Wash. U contlig 1003.  
Seq primer: TR

Class: sheared ends.

FEATURES  
source Location/Qualifiers  
1..199

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="AUHHC32"

/clone.lib="AUH"

/note="Vector: PHOS2; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into PHOS2 using BstXI linkers"

BASE COUNT 55 a 42 c 34 g 68 t  
ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 199;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCGCTGCCAATTG 18  
|||||  
DB 154 TGGCGCGCTGCCAATTG 171

RESULT 6

CNS00T11/c 216 bp DNA linear GSS 28-JUN-1999  
LOCUS Arabidopsis thaliana genome survey sequence T7 end of BAC T3023 of

DEFINITION TAMU library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.

ACCESSION AT089107  
VERSION AT089107.1 GI:5290247  
KEYWORDS GSS.

SOURCE Arabidopsis thaliana.  
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 216)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

AUTHORS 1 (bases 1 to 216)  
Samsou, D., Saurin, W., Choise, N., Artiguenave, F., Brotlier, P., Wincker, P.,

Salanoubat, M., Choisne, N., Artiguenave, F., Brotlier, P., Wincker, P.,

Salanoubat, M., Saurin, W., Weissenbach, J. and Quetier, F.

Unpublished

2 (bases 1 to 216)

Genoscope.

Direct Submission

Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr))

- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

FEATURES  
source Location/Qualifiers  
1..216

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

```

/clone="T3D23"
/clone.lib="TAMU"
/note="end : T7"
BASE COUNT      82 a      40 c      40 g      54 t
ORIGIN
Query Match      100.0%; Score 18; DB 17; Length 216;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY      1 TGCGCGCGGTGCCAATTG 18
      |||||||
Db      135 TGCGCGCGGTGCCAATTG 118
      |||||||

RESULT 7
BH243803      263 bp      DNA      linear      GSS 13-NOV-2001
DEFINITION    AU1R1B17R AU1R Arabidopsis thaliana genomic clone AU1R17, DNA
sequence.
ACCESSION     BH243803
VERSION       BH243803.1 GI:16920146
KEYWORDS      GSS.
SOURCE        thale cress.
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 263)
Town,C.D., Whitelaw,C.A., Pal,G., Van Aken,S.E., Utterback,T.V.,
Feldblyum,T.V. and Fraser,C.M.,
Survey sequencing of Arabidopsis thaliana BAC F9022
Unpublished (2001)
Other-GSSs: AU1R17TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contlg 1257.
Seq primer: TR
Class: sheared ends.
location/Qualifiers
1. 263
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AU1R17"
/clone.lib="AU1R"
/note="Vector: pHO52; Site.1: BstXI; 2-3 kb sheared BAC
DNA inserted into pHO52 using BstXI linkers"
BASE COUNT      61 a      45 c      48 g      109 t
ORIGIN
Query Match      100.0%; Score 18; DB 17; Length 263;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 TGCGCGCGGTGCCAATTG 18
      |||||||
Db      136 TGCGCGCGGTGCCAATTG 153
      |||||||

RESULT 8
BH2585/c
LOCUS         B62585/c      388 bp      DNA      linear      GSS 21-NOV-1997
DEFINITION    T22F21TF TAMU Arabidopsis thaliana genomic clone T22F21, DNA
sequence.
ACCESSION     B62585
VERSION       B62585.1 GI:2629347
KEYWORDS      GSS.
SOURCE        thale cress.

```

```

ORGANISM
  Arabidopsis thaliana
  Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
  Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsis.
  1 (bases 1 to 388)
  Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.,
  Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
  ,J.C.
  A BAC End Sequence database for identifying Minimal Overlaps in
  Arabidopsis Genomic Sequencing. Update 3
  Unpublished (1997)
  Contact: Steve Rounsley
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: rounsley@tigr.org
  Seq primer: ML3-21
  Class: BAC ends
  High quality sequence stop: 388.

FEATURES
  source
    1..388
    /organism="Arabidopsis thaliana"
    /strain="Columbia"
    /db_xref="taxon:3702"
    /clone="T22F21"
    /clone_1lb="TAMU"
    /sex="hermaphrodite"
    /note="Vector: BelobAC11; Site_1: HindIII; Site_2: HindIII
    ; Produced by Rod Wing"
  BASE COUNT
    143 a 71 c 77 g 97 t
  ORIGIN
    Query Match
    Best Local Similarity 100.0%; Score 18; DB 17; Length 388;
    Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    QY
    1 TGGCGCCGTTGCCAATTG 18
    |||||||
    DB
    142 TGGCGCCGTTGCCAATTG 125
    |||||||

RESULT 9
BH244138
LOCUS
  BH244138 401 bp DNA linear GSS 13-NOV-2001
  AUTMA26TF AUTM Arabidopsis thaliana genomic clone AUTMA26, DNA
  sequence.
ACCESSION
  BH244138
VERSION
  BH244138.1 GI:16920790
KEYWORDS
  GSS.
SOURCE
  thale cress.
  Arabidopsis thaliana.
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsis.
  1 (bases 1 to 401)
  Town,C.D., Whitelaw,C.A., Pal,G., Van Aken,S.E., Uterback,T.V.,
  Feldblyum,T.V. and Fraser,C.M.
  Survey sequencing of Arabidopsis thaliana BAC F13C23
  Unpublished (2001)
  Other GSSs: AUTMA26TR
  Contact: Chris Town
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA.
  Tel: 301-838-3523
  Fax: 301-838-0208
  Email: cdtown@tigr.org
  Gen Wash. U contig 1440.
  Seq primer: TP
  Class: sheared ends.
  Location/Qualifiers
    1..401
  FEATURES
  source
    1..401
  Location/Qualifiers

```

```

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AUTWA26"
/clone_11b="AUTW"
/Note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
DNA inserted into PHOS2 using BstXI linkers"
BASE COUNT      114 a      91 c      69 g      127 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 401;
Best Local Similarity 100.0%; Pred. NO. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCGGTGCCAATTG 18
DB      118 TGGCGCGGTGCCAATTG 135

RESULT 10
BH244101/c      410 bp      DNA      linear      GSS 13-NOV-2001
LOCUS      AUTWA38TR AUTW Arabidopsis thaliana genomic clone AUTWA38, DNA
DEFINITION
ACCESSION      BH244101
VERSION
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 410)
Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Uterback,T.V.,
Feidblyum,T.V. and Fraser,C.M.
Survey sequencing of Arabidopsis thaliana BAC F13C23
Unpublished (2001)
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1440.
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..410
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AUTWA38"
/clone_11b="AUTW"
/Note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
DNA inserted into PHOS2 using BstXI linkers"
BASE COUNT      125 a      70 c      93 g      122 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 410;
Best Local Similarity 100.0%; Pred. NO. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCGGTGCCAATTG 18
DB      302 TGGCGCGGTGCCAATTG 285

RESULT 11
BH243289
LOCUS      AUTHF63TR AUTH Arabidopsis thaliana genomic clone AUTHF63, DNA
DEFINITION
ACCESSION      BH243289

```

```

VERSION      BH243289.1 GI:16919186
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 415)
Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Uterback,T.V.,
Feidblyum,T.V. and Fraser,C.M.
Survey sequencing of Arabidopsis thaliana BAC F1307
Unpublished (2001)
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1003.
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..415
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AUTHF63"
/clone_11b="AUTH"
/Note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
DNA inserted into PHOS2 using BstXI linkers"
BASE COUNT      116 a      99 c      81 g      119 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 415;
Best Local Similarity 100.0%; Pred. NO. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCGGTGCCAATTG 18
DB      8 TGGCGCGGTGCCAATTG 25

RESULT 12
BH243156/c      418 bp      DNA      linear      GSS 13-NOV-2001
LOCUS      AUTHC93TR AUTH Arabidopsis thaliana genomic clone AUTHC93, DNA
DEFINITION
ACCESSION      BH243156
VERSION
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 418)
Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Uterback,T.V.,
Feidblyum,T.V. and Fraser,C.M.
Survey sequencing of Arabidopsis thaliana BAC F1307
Unpublished (2001)
Other GSSs: AUTHC93TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1003.
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..418
/organism="Arabidopsis thaliana"

```

/strain="Columbia"  
/db.xref="taxon:3702"  
/clone="AUIHC93"  
/clone\_1lb="AUIH"  
/note="Vector: PHOS2; Site\_1: BstXI; 2-3 kb sheared BAC  
DNA inserted into PHOS2 using BstXI linkers"

BASE COUNT

152 a 82 c 94 g 90 t

Query Match

Best Local Similarity 100.0%; Score 18; DB 17; Length 418;  
Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18

Db 34 TGGCGCCGTTGCCAATTG 17

RESULT 13

BH252783

LOCUS BH252783 419 bp DNA linear GSS 28-NOV-2001  
DEFINITION SALK\_013862 Arabidopsis thaliana TDNA insertion lines Arabidopsis  
thaliana genomic clone SALK\_013862, DNA sequence.

ACCESSION

BH252783

VERSION BH252783.1 GI:17139761

KEYWORDS

SOURCE

thale cress.

ORGANISM

REFERENCE

AUTHORS

Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,  
C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,  
Zimmerman,J. and Ecker,J.R.  
A sequence-indexed library of insertion mutations in the  
Arabidopsis genome  
Unpublished (2001)

TITLE

JOURNAL

COMMENT

Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGNAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: eckersalk.edu  
This is single pass sequence recovered from the left border of  
TDNA.  
Class: TDNA tagged.

FEATURES

SOURCE

Location/Qualifiers  
1..419  
/organism="Arabidopsis thaliana"  
/strain="Columbia 0"  
/db.xref="taxon:3702"  
/clone="SALK\_013862"  
/clone\_1lb="Arabidopsis thaliana TDNA insertion lines"  
/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocol used can  
be found at [http://signal.salk.edu/cdna\\_protocols.html](http://signal.salk.edu/cdna_protocols.html)"

BASE COUNT 105 a 99 c 146 g 146 t

Query Match

Best Local Similarity 100.0%; Score 18; DB 17; Length 419;  
Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18

Db 209 TGGCGCCGTTGCCAATTG 226

RESULT 14

BH244105/c

LOCUS

BH244105 422 bp DNA linear GSS 13-NOV-2001  
DEFINITION AUIWA21TR AUIW Arabidopsis thaliana genomic clone AUIWA21, DNA  
sequence.

ACCESSION

BH244105

VERSION

BH244105.1 GI:16920722

KEYWORDS

SOURCE

thale cress.

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
From Wash. U contig 1440.  
Seq primer: TR  
Class: sheared ends.

FEATURES

Location/Qualifiers  
1..422  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db.xref="taxon:3702"  
/clone="AUIWA21"  
/clone\_1lb="AUIW"  
/note="Vector: PHOS2; Site\_1: BstXI; 2-3 kb sheared BAC  
DNA inserted into PHOS2 using BstXI linkers"

BASE COUNT 160 a 81 c 86 g 95 t

Query Match

Best Local Similarity 100.0%; Score 18; DB 17; Length 422;  
Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18

Db 96 TGGCGCCGTTGCCAATTG 79

RESULT 15

BH244079

LOCUS

BH244079 424 bp DNA linear GSS 13-NOV-2001  
DEFINITION AUIRB53TR AUIR Arabidopsis thaliana genomic clone AUIRB53, DNA  
sequence.

ACCESSION

BH244079

VERSION

BH244079.1 GI:16920670

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org

From Wash. U contig 1257.

Seq primer: TR

Class: sheared ends.  
Location/Qualifiers

FEATURES  
Source  
1. 424  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="AU1RB53"  
/clone\_lib="AU1R"  
/note="Vector: PHOS2; Site\_1: BstXI; 2-3 kb sheared BAC  
DNA inserted into PHOS2 using BstXI linkers"  
BASE COUNT 94 a 83 c 83 g 164 t  
ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 424;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
Db 324 TGGCGCCGTTGCCAATTG 341

Search completed: June 20, 2003, 22:39:12  
Job time : 71.1132 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 19:31:16 ; Search time 75.6226 Seconds  
(without alignments)  
6927.165 Million cell updates/sec

Title: US-09-965-553-1

Sequence: 1 tggcgccttgccatg 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

```

1: GenEmbl:
2: gb_ba:
3: gb_hg:
4: gb_in:
5: gb_om:
6: gb_ov:
7: gb_pat:
8: gb_ph:
9: gb_pl:
10: gb_pr:
11: gb_ro:
12: gb_scs:
13: gb_sy:
14: gb_un:
15: gb_vi:
16: em_ba:
17: em_fun:
18: em_hum:
19: em_in:
20: em_mu:
21: em_om:
22: em_or:
23: em_ov:
24: em_ph:
25: em_pl:
26: em_ro:
27: em_sts:
28: em_un:
29: em_vi:
30: em_hg_hum:
31: em_hg_inv:
32: em_hg_other:
33: em_hg_mus:
34: em_hg_dln:
35: em_hg_tod:
36: em_hg_mam:
37: em_hg_vrt:
38: em_sy:
39: em_hggo_hum:
40: em_hggo_mus:
41: em_hggo_other:

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	100.0	3539	8	AB073161	AB073161 Arabidops
2	18	100.0	18433	8	AC093090	AC093090 Arabidops
3	18	100.0	18813	8	T9E19	AF104920 Arabidops
4	18	100.0	32259	2	AC109921	AC109921 Arabidops
5	18	100.0	33329	8	AC007261	AC007261 Arabidops
6	18	100.0	35551	6	AX059461	AX059461 Sequence
7	18	100.0	36032	6	AX059500	AX059500 Sequence
8	18	100.0	38519	6	AX059548	AX059548 Sequence
9	18	100.0	39104	6	AX059479	AX059479 Sequence
10	18	100.0	40480	6	AX059456	AX059456 Sequence
11	18	100.0	42112	6	AX059497	AX059497 Sequence
12	18	100.0	42208	6	AX059477	AX059477 Sequence
13	18	100.0	47383	6	AX059459	AX059459 Sequence
14	18	100.0	47840	6	AX059452	AX059452 Sequence
15	18	100.0	48128	6	AX059469	AX059469 Sequence
16	18	100.0	48128	6	AX059474	AX059474 Sequence
17	18	100.0	48128	6	AX059513	AX059513 Sequence
18	18	100.0	48422	6	AX059509	AX059509 Sequence
19	18	100.0	50349	8	AC079028	AC079028 Arabidops
20	18	100.0	52616	2	AC090029	AC090029 Arabidops
21	18	100.0	54573	8	54573	AF147265 Arabidops
22	18	100.0	55345	8	AC007209	AC007209 Arabidops
23	18	100.0	57000	8	AB046428	AB046428 Arabidops
24	18	100.0	57000	8	AB046428	AB046428 Arabidops
25	18	100.0	58411	8	AB046430	AB046430 Arabidops
26	18	100.0	58411	8	AB046430	AB046430 Arabidops
27	18	100.0	60482	2	AC090030	AC090030 Arabidops
28	18	100.0	62916	8	AB046427	AB046427 Arabidops
29	18	100.0	64165	2	AC067965	AC067965 Arabidops
30	18	100.0	65316	8	FLAC23	AF296828 Arabidops
31	18	100.0	66542	8	T10118	AF262040 Arabidops
32	18	100.0	68352	8	T15F17	AF262042 Arabidops
33	18	100.0	69752	8	T5H22	AF096372 Arabidops
34	18	100.0	74265	8	AC068901	AC068901 Arabidops
35	18	100.0	76303	2	AC007183	AC007183 Arabidops
36	18	100.0	76911	2	AC016828	AC016828 Arabidops
37	18	100.0	77287	8	AB026642	AB026642 Arabidops
38	18	100.0	80279	2	AC011624	AC011624 Arabidops
39	18	100.0	81414	8	AB024037	AB024037 Arabidops
40	18	100.0	81806	8	AB062087	AB062087 Arabidops
41	18	100.0	81902	8	T17A2	AF160183 Arabidops
42	18	100.0	81902	8	T17A2	AF160183 Arabidops
43	18	100.0	83129	2	AC009529	AC009529 Arabidops
44	18	100.0	83530	8	AB062093	AB062093 Arabidops
45	18	100.0	84711	8	AP002067	AP002067 Arabidops

#### ALIGNMENTS

```

RESULT 1
AB073161/c 3539 bp DNA linear PLN 26-FEB-2002
LOCUS Arabidopsis thaliana DNA, chromosome 4 centromere region, BAC
DEFINITION
ACCESSION AB073161
VERSION AB073161
KEYWORDS Arabidopsis thaliana
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana (cultivar:Columbia) DNA, clone:T5110.
REFERENCE 1
AUTHORS Kumeiawa,N., Hosouchi,T., Tsuruoka,H. and Kotani,H.

```

**TITLE** The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 4

**JOURNAL** DNA Res 8 (6), 285-290 (2001)

**REFERENCE** 2 (bases 1 to 3539)

**AUTHORS** Kotani, H. and KumeKawa, N.

**JOURNAL** Direct Submission

Submitted (16-OCT-2001) Hirokazu Kotani, Kazusa DNA Research Institute, Lab. Chromosome Research II, 1532-3 Yana, Kasarazu, Chiba 292-0812, Japan (E-mail: kotani@kazusa.or.jp, Tel:81-438-52-3920, Fax:81-438-52-3921)

**FEATURES**

**Source**

1. 3539  
 Location/Qualifiers  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="4"  
 /map="centromere"  
 /clone="T5110"  
 /note="this sequence was derived by PCR to cover the gap between clone T11G11 and clone F8H12"  
 1. 1000  
 /note="overlap to BAC T11G11, please refer for analysis and annotation"  
 misc\_feature  
 1001..2539  
 /note="similarity to fragments of Athlia retrotransposons"  
 2540..3539  
 /note="overlap to BAC F8H12, please refer for analysis and annotation"  
 misc\_feature  
 1025 a 734 c 815 g 965 t  
 BASE COUNT  
 ORIGIN

**Query Match** 100.0%; Score 18; DB 8; Length 3539;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Db** 1 TGGCGCCGTTGCCAATTG 18  
 1702 TGGCGCCGTTGCCAATTG 1685

**RESULT 2**  
**AC093090** 18433 bp DNA linear PLN 31-AUG-2001  
**LOCUS** Arabidopsis thaliana chromosome 3 BAC F1C23 genomic sequence,  
**DEFINITION** complete sequence.  
**AC093090**  
**VERSION** AC093090.1 GI:15145156  
**KEYWORDS** HTG.  
**SOURCE** Arabidopsis thaliana.  
**ORGANISM** Arabidopsis thaliana.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 18433)  
 Town, C.D., Haas, B.J., Wu, D., Maitl, R., Hanick, L.T., Chan, A.P., Tallon, L.J., Rooney, T., Utterback, T.R., Vanaken, S.E., Feldlyum, T.V., White, O. and Fraser, C.M.  
 Arabidopsis thaliana chromosome 3 BAC F1C23 genomic sequence  
 Unpublished  
 2 (bases 1 to 18433)  
 Town, C.D., White, O. and Fraser, C.M.  
**REFERENCE** Direct Submission  
**AUTHORS** Submitted (09-AUG-2001) The Institute for Genomic Research, 9712  
**JOURNAL** Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org  
 3 (bases 1 to 18433)  
 Town, C.D. and Kaul, S.  
**REFERENCE** Direct Submission  
**AUTHORS** Submitted (31-AUG-2001) The Institute for Genomic Research, 9712  
**JOURNAL** Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org  
**COMMENT** Address all correspondence to: atetlgr.org  
 BAC clone F1C23 is from Arabidopsis thaliana chromosome 3

The orientation of the sequence is from SP6 to T7 end of the BAC clone.  
 Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mt.edu/GENSCAN.html>), GenemarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/Genemark/>), Glimmer4 (a variant of Glimmer3, see Minaela Perlea, <http://www.tigr.org/softlab/glimmer4.htm> and Geneslicer (Minaela Perlea and Steven Salzberg, <http://mergatel.tigr.org>), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arvan Smil, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).  
 Location/Qualifiers  
 1. 18433  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="3"  
 /map="?"  
 /clone="F1C23"  
 complement(365..405)  
 /rpt\_family="AT-rich"  
 2852..5536  
 /gene="F1C23.1"  
 /note="pseudogene, putative Athlia ORF1"  
 /pseudo  
 complement(16225..16754)  
 /gene="F1C23.2"  
 /note="similar to protein prospero GI:1346808 (Drosophila melanogaster)"  
 complement(16225..16384,16614..>16754))  
 /gene="F1C23.2"  
 complement(join(16226..16384,16614..16754))  
 /gene="F1C23.2"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAK96023.1"  
 /db\_xref="GI:15383840"  
 /translation="MCMAYLVVFRYGCWSSNRGTEHEVRMPDPAFSKTSLOPWS PARGRWTIOSRHHSYISLDKRTPLKPSLINDIDHWKAHFSPLDANDTGRRLMT"  
 BASE COUNT  
 ORIGIN  
 5393 a 3632 c 3666 g 5742 t

**Query Match** 100.0%; Score 18; DB 8; Length 18433;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Db** 1 TGGCGCCGTTGCCAATTG 18  
 2563 TGGCGCCGTTGCCAATTG 2580

**RESULT 3**  
**T9E19/C** 18813 bp DNA linear PLN 11-NOV-1998  
**LOCUS** Arabidopsis thaliana BAC T9E19.  
**DEFINITION** Arabidopsis thaliana BAC T9E19.  
**AC093090** AF104920  
**VERSION** AF104920.1 GI:3859610  
**KEYWORDS** Arabidopsis thaliana.  
**SOURCE** Arabidopsis thaliana.  
**ORGANISM** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;



REFERENCE Rosidae: eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 18813)  
AUTHORS Washington University Genome Sequencing Center.  
TITLE The A. thaliana Genome Sequencing Project.  
JOURNAL Unpublished (1997)  
REFERENCE 2 (bases 1 to 18813)  
AUTHORS Stromatt,C., Johnson,D. and Le,T.  
TITLE The sequence of A. thaliana T9E19  
JOURNAL Unpublished (1998)  
REFERENCE 3 (bases 1 to 18813)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (08-NOV-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63108, USA  
e-mail: twilson@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

#### NEIGHBORING COSMID INFORMATION:

The 5' clone is T2L5, 200 bp overlap; 3' clone is F7N22, 200 bp overlap. Actual start of this clone could not be found, the first known base of overlap is at 36622 of T2L5; actual end is at 76001 of F7N22.

#### NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hallier, ms in preparation).

#### FEATURES

##### SOURCE

1. 18813  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="IV"  
/map="5"  
/clone="T9E19"  
57. 1256  
/gene="T9E19.2"  
join(57. 893, 969. 1256)  
/gene="T9E19.2"  
/codon\_start=1  
/evidence-not\_experimental  
/protein\_id="AAC72876.1"  
/db\_xref="gi:3859611"  
/translation="MNAKSPKPEYISDIENVLKGPMPFLDPSFGELEKIPONKASF  
NAKVLGLICRLQIVRKKNEMIVFGGHPHREGIREFSITLIGEGKYPKKKVDVVI  
NVKPECSVMNTLDERGCDNYPTIADYSMTQEPESMGWKLALSLIIVGVYAA  
HSNRPRTSKYVEKTKLKFCKTIPMGKYSFTRTGLRINAFQIPYAOQLIRLLVGS  
YALGFLPLQLDLAFETIPSLAKLGPDDVILNTPFARSLHRLASLAIKRIILCEEA  
ADEEVANIVKPADVNCPSLSMDDEVDPKVDYIALIDGQWODEMDEWVGAYARP  
KQVRPQLEETGKYKKGCIQIFLVLHPEETQLIATGILL"

##### gene

##### CDS

gene 12757. 15124  
/gene="T9E19.1"  
join(12757. 13350, 13635. 13799, 13898. 14023, 14096. 14200,  
14290. 14391, 14462. 14521, 14645. 15124)  
/gene="T9E19.1"  
/codon\_start=1  
/evidence-not\_experimental  
/protein\_id="AAC72877.1"  
/db\_xref="gi:3859612"

##### CDS

BASE COUNT 5547 a 3923 c 3965 g 5378 t  
ORIGIN

Query Match 100.0%; Score 18; DB 8; Length 18813;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGCGCCGTTGCCAATTG 18  
Db 11028 TGCGCCGTTGCCAATTG 11011

#### RESULT 4

##### AC109921

##### LOCUS

##### DEFINITION

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

##### TITLE</

```

*      15084      24483: contig of 9400 bp in length
*      24484      24524: gap of unknown length
*      24525      32259: contig of 7735 bp in length.
FEATURES
    source
        1..32259
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /chromosome="unknown"
            /clone="IGF-F9022"
BASE COUNT      9587 a      6125 c      6878 g      9398 t      271 others
ORIGIN
Query Match      100.0%; Score 18; DB 2; Length 32259;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 TGGCCCGCTGCCAATTG 18
        |||
        16753 TGGCCCGCTGCCAATTG 16770

RESULT 5
AC007261      33329 bp      DNA      linear      PLN 27-FEB-2002
LOCUS      Arabidopsis thaliana chromosome 2 clone F16G22 map g4532, complete
DEFINITION
AC007261
VERSION      AC007261.5 GI:20198225
KEYWORDS      HMG.
SOURCE      Arabidopsis thaliana.
ORGANISM      Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 33329)
Lin, X., Kaul, S., Shea, T. P., Fujii, C. Y., Shen, M., VanAken, S. E.,
Barnstead, M. E., Mason, T. M., Bowman, C. L., Ronning, C. M.,
Bellio, M. I., Carrera, A. J., Creasy, T. H., Buell, C. R., Town, C. D.,
Merriman, W. C., Fraser, C. M. and Venter, J. C.
Unpublished
2 (bases 1 to 33329)
Lin, X.
Direct Submission
JOURNAL      Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
REFERENCE      Medical Center Dr., Rockville, MD 20850, USA
AUTHORS      Town, C.D. and Kaul, S.
JOURNAL      Direct Submission
COMMENT      Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
FEATURES      Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org
                On Apr 18, 2002 this sequence version replaced gi:6598734.
                Location/Qualifiers
                source
                    1..33329
                        /organism="Arabidopsis thaliana"
                        /cultivar="Columbia"
                        /db_xref="taxon:3702"
                        /chromosome="2"
                        /map="g4532"
                        /clone="F16G22"
                        /complement(1..482)
                        /note="overlap with BAC clone F15K19 (AC006429:1..482)."
                    misc_feature
                        1
                        /note="2721 nt before this point were not included in the
                        submitted sequence, due to overlap with another BAC"
                    repeat_region
                        /rpt_family="AT_rich"
                        complement(383..431)
                    repeat_region
                        /rpt_family="AT_rich"
                        complement(2109..2135)
                    repeat_region
                        /rpt_family="AT_rich"
                        complement(2211..2267)
                    repeat_region
                        /rpt_family="AT_rich"
                        2482..2582
                    repeat_region
                        /rpt_family="AT_rich"

```

```

repeat_region      2685..2724
                    /rpt_family="AT_rich"
repeat_region      2951..2992
                    /rpt_family="AT_rich"
repeat_region      complement(3184..3205)
                    /rpt_family="AT_rich"
gene              5423..6109
                    /gene="At2g10930"
                    /note="F16G22.1"
                    join(5423..5657,5687..5861,5984..6109)
                    /gene="At2g10930"
                    join(5472..5657,5687..5861,5984..6072)
                    /gene="At2g10930"
                    /codon_start=1
                    /product="unknown protein"
                    /protein_id="AAD28645.1"
                    /db_xref="GI:4733962"
                    /translation="MEVSDSSQVLYNSKKHVASLRADDEPDESLISSTLIDLVFR
                    LVLYGVVIMLDQVQETALSGEVNMEGSRREFNPMPKPIVYKRMKAPKHYDE
                    DLKIFPQATTYILKIQKGGSSGCFEIRGEVCTKDSQNNTFYIF"
                    complement(6054..6076)
                    /rpt_family="AT_rich"
                    6215..7085
                    /note="DNA sequence similar to L1mpe1 transposon
                    GB:U76697"
                    7256..7507
                    /gene="At2g10920"
                    /note="F16G22.2"
                    7256..7507
                    /gene="At2g10920"
                    7271..7442
                    /rpt_family="(TGG)n"
                    7310..7507
                    /gene="At2g10920"
                    /codon_start=1
                    /product="unknown protein"
                    /protein_id="AAD28646.1"
                    /db_xref="GI:4733963"
                    /translation="MNTFEVVAEMEAIVITRDIAEYVEKDIYVAVREVAEEVTL
                    EAREIEBMGLIVRDSLYLCRR"
                    complement(7921..7948)
                    /rpt_family="AT_rich"
                    complement(8517..13115)
                    /gene="At2g10910"
                    /note="F16G22.3: pseudogene, Tail-1-like non-LTR
                    retroelement reverse transcriptase"
                    /pseudo
                    complement(11206..11773)
                    /gene="At2g10910"
                    /note="DNA sequence similar to non-LTR retrotransposon
                    pseudogene for reverse transcriptase GB:LA7192"
                    /pseudo
                    complement(13353..13378)
                    /rpt_family="AT_rich"
                    complement(13763..16746)
                    /gene="At2g10900"
                    /note="F16G22.4: similar to GB:AAA75253; some members of
                    this protein family have a weak CCHC zinc fingers that is
                    mostly from retroviral gag proteins (nucleocapsid)"
                    complement(join<13763..14997,15212..15445,15881..16084,
                    16536..>16746))
                    /gene="At2g10900"
                    complement(join<13763..14997,15212..15445,15881..16084,
                    16536..16746))
                    /gene="At2g10900"
                    /codon_start=1
                    /product="putative Tail-1-like non-LTR retroelement protein"
                    /protein_id="AAD28647.1"
                    /db_xref="GI:4733964"
                    /translation="MMVITGVHGVIGVNDGSGYSTKSSGVYKSGSRLLDLY
                    EKWIRCGRVTCRCGCVYQZLODLDLFDLKSSSVYFTYIQGDFLSFVRLPOT
                    RYVYVLEKGRSGVILIDAPGRFVSGSEAKQIVIGGRMSQSGLVNQSNSVNTMKL
                    KIVRFNNSLIIEGYSKTLIGRCMNPADMKSLFMLPKIMKLEDRVYGVANLGGGE

```

```

VOAVDLDFGRVQVNLDFGFKPLCFEAAVEFHSGSETMWSLWYERLFGFCORCYSICHEO
MRCPIFGDKGKOSJFEKEDPKRDNTLQKRGALINGSPGTSSTGSAVAEAPGOLTS
IDKDDASIODSYNSKOSIGIKAKGRNLLFVAPVQASSSOMAVOYTKLSQLOTDTG
MGEPDHEKKMLFARLAPRAGKTPTDGLTGAMEGILSTSPSKYRRKFFPSGPEYV
IONKANLELAEQHDLEFDDLSQLAQSPLGDPSTQFMDTKMEDKSSGVELTHE
DGEFFAANEENFDGKLEVLPLPINTVSEMNVMVSVSENEQAHHEEKSQGE
DELETDPAKRRMKRNRKFLTVSVSSRRKNYLTSPRRKVPKPGCTPVGAGSHQREKRP
IKH"
/misc_feature
16595..25479
/feature="putative retroelement. Structure: LTR,
retroelement, LTRB"
retroelement(17799..17858)
complement(17799..17858)
/rpt_family="Rf:centromeric repeat from T7M24 17342 to
19092 1751nt includes within it 106B centromeric repeat"
18341..21479
/gene="At2g10890"
/feature="F16522.5; pseudogene, Athlia retroelement ORP1"
/pseudo
22177..22210
/rpt_family="AT-rich"
22349..22631
/rpt_family="Rf:ATHILA/ATHILA A.thaliana DNA for Athlia
retroelement.#Rf:ATRO033|X81801 Athlia retroelement"
22416..22595
/rpt_family="Rf:q1|895701|emb|X89195.1|ATSATDVA1
A.thaliana DNA of a 180bp satellite
junction#Rf:ATRO034|X89195 clone H12 of Athlia 180-bp
repeat"
22629..22790
/rpt_family="Rf:ATHILA/ATHILA A.thaliana DNA for Athlia
retroelement.#Rf:ATRO033|X81801 Athlia retroelement"
22629..22733
/rpt_family="Rf:q1|895701|emb|X89195.1|ATSATDVA1
A.thaliana DNA of a 180bp satellite
junction#Rf:ATRO034|X89195 clone H12 of Athlia 180-bp
repeat"
23055..23229
/rpt_family="Rf:ATHILA/ATHILA A.thaliana DNA for Athlia
retroelement.#Rf:ATRO033|X81801 Athlia retroelement"
24047..24310
/gene="At2g10880"
/feature="F16522.6; similar to hypothetical protein
GB:AC26673"
<24047..24310
/gene="At2g10880"
24047..24310
/gene="At2g10880"
/codon_start=1
/product="hypothetical protein"
/db_xref="GI:4733965"
/translation="MAEYHHHSLDQSLYHOVEYNNHHPLDQSLYHVESLIHIST
LIDQQLSLHSLNQTLEHKOEKTPAYHSTHSTWYERS"
complement(24277..24387)
/rpt_family="Rf:centromeric repeat from T7M24 17342 to
19092 1751nt includes within it 106B centromeric repeat"
24643..24711
/rpt_family="Rf:106B centromeric repeat X33611, 106B is a
diverged copy of the Athlia retroelement, 300
copies#Rf:911100798|emb|X93611.1|AT106B A.thaliana
centromeric repeat region (clone 106B)#Rf:ATRO044|X93611
106B"
complement(25075..25134)
/rpt_family="Rf:centromeric repeat from T7M24 17342 to
19092 1751nt includes within it 106B centromeric repeat"
complement(25459..25710)
/gene="At2g10870"
/feature="F16522.7; predicted by genefinder"
complement(25459..25710)
/gene="At2g10870"
complement(25459..25710)
/codon_start=1

```

```

/product="hypothetical protein"
/protein_id="AAD28649.1"
/db_xref="GI:4733965"
/translation="MEFDGVTALPAVVAEAFVVEALYSORLKRKRRLREBEE
DDEPDLGDANNEEDCGVYGDGDCVAVEDIVGGGENDN"
Query Match
Best Local Similarity 100.0%; Score 18; DB 8; Length 33329;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCGCGCGTGCATTG 18
|||||
Db 18199 TGCGCGCGTGCATTG 18216
|||||
RESULT 6
AX059461/c
LOCUS AX059461 35551 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 194 from Patent WO0055325.
ACCESSION AX059461
VERSION AX059461.1 GI:12311566
KEYWORDS
SOURCE
Arabidopsis thaliana
thalae cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 35551)
Preuss,D., Copenhaver,G. and Keith,K.
Plant chromosome compositions and methods
Patent: WO 0055325-A 194 21-SEP-2000;
The University of Chicago (US)
Location/Qualifiers
1..35551
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
BASE COUNT 10624 a 7544 c 7078 g 10298 t 7 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 18; DB 6; Length 35551;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCGCGCGTGCATTG 18
|||||
Db 23423 TGCGCGCGTGCATTG 23406
|||||
RESULT 7
AX059500
LOCUS AX059500 36032 bp DNA linear PAT 18-JAN-2001
DEFINITION Sequence 233 from Patent WO0055325.
ACCESSION AX059500
VERSION AX059500.1 GI:12311607
KEYWORDS
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 36032)
Preuss,D., Copenhaver,G. and Keith,K.
Plant chromosome compositions and methods
Patent: WO 0055325-A 233 21-SEP-2000;
The University of Chicago (US)
Location/Qualifiers
1..36032
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
BASE COUNT 10923 a 7401 c 7909 g 9799 t
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 36032;

```

Best Local Similarity 100.0%; Pred. No. 79;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
Db 5453 TGGCGCCGTTGCCAATTG 5470

RESULT 8  
AX059548  
LOCUS AX059548 38519 bp DNA linear PAT 17-JAN-2001  
DEFINITION Sequence 281 from Patent WO0055325.  
ACCESSION AX059548  
VERSION AX059548.1 GI:12311655  
KEYWORDS  
SOURCE  
ORGANISM thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 38519)  
AUTHORS Preuss, D., Copenhagen, G. and Keith, K.  
TITLE Plant chromosome compositions and methods  
JOURNAL Patent: WO 0055325-A 281 21-SEP-2000;  
The University of Chicago (US)  
FEATURES  
Location/Qualifiers  
1..38519  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"

BASE COUNT 11817 a 7622 c 7434 g 11646 t  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 38519;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
Db 238 TGGCGCCGTTGCCAATTG 255

RESULT 9  
AX059479 39104 bp DNA linear PAT 17-JAN-2001  
LOCUS AX059479  
DEFINITION Sequence 212 from Patent WO0055325.  
ACCESSION AX059479  
VERSION AX059479.1 GI:12311584  
KEYWORDS  
SOURCE thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 39104)  
AUTHORS Preuss, D., Copenhagen, G. and Keith, K.  
TITLE Plant chromosome compositions and methods  
JOURNAL Patent: WO 0055325-A 212 21-SEP-2000;  
The University of Chicago (US)  
FEATURES  
Location/Qualifiers  
1..39104  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"

BASE COUNT 11558 a 7680 c 7833 g 12033 t  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 39104;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
Db 7762 TGGCGCCGTTGCCAATTG 7779

RESULT 10  
AX059456/c 40480 bp DNA linear PAT 17-JAN-2001  
LOCUS AX059456  
DEFINITION Sequence 189 from Patent WO0055325.  
ACCESSION AX059456  
VERSION AX059456.1 GI:12311561  
KEYWORDS  
SOURCE thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 40480)  
AUTHORS Preuss, D., Copenhagen, G. and Keith, K.  
TITLE Plant chromosome compositions and methods  
JOURNAL Patent: WO 0055325-A 189 21-SEP-2000;  
The University of Chicago (US)  
FEATURES  
Location/Qualifiers  
1..40480  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"

BASE COUNT 12306 a 7912 c 8765 g 11297 t 200 others  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 40480;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
Db 35456 TGGCGCCGTTGCCAATTG 35439

RESULT 11  
AX059497 42112 bp DNA linear PAT 17-JAN-2001  
LOCUS AX059497  
DEFINITION Sequence 230 from Patent WO0055325.  
ACCESSION AX059497  
VERSION AX059497.1 GI:12311604  
KEYWORDS  
SOURCE thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 42112)  
AUTHORS Preuss, D., Copenhagen, G. and Keith, K.  
TITLE Plant chromosome compositions and methods  
JOURNAL Patent: WO 0055325-A 230 21-SEP-2000;  
The University of Chicago (US)  
FEATURES  
Location/Qualifiers  
1..42112  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"

BASE COUNT 12877 a 8525 c 8423 g 12287 t  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 42112;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
Db 1458 TGGCGCCGTTGCCAATTG 1441

RESULT 12  
AX059477/c 42208 bp DNA linear PAT 17-JAN-2001  
LOCUS AX059477  
DEFINITION Sequence 210 from Patent WO0055325.  
ACCESSION AX059477  
VERSION AX059477.1 GI:12311582

KEYWORDS  
SOURCE  
ORGANISM  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1 (bases 1 to 42208)  
Preuss,D., Copenhaver,G. and Keith,K.  
Plant chromosome compositions and methods  
Patent: WO 005325-A 210 21-SEP-2000;  
The University of Chicago (US)  
FEATURES  
source  
1.42208  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
BASE COUNT 12522 a 8859 c 886 g 11961 t  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 42208;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG.18  
|||||  
Db 35889 TGGCGCCGTTGCCAATTG 35872

RESULT 13  
AX059459/c 47383 bp DNA linear PAT 17-JAN-2001  
LOCUS  
DEFINITION  
Sequence 192 from Patent WO0055325.  
AX059459  
ACCESSION  
VERSION  
AX059459.1 GI:12311564  
KEYWORDS  
SOURCE  
ORGANISM  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1 (bases 1 to 47383)  
Preuss,D., Copenhaver,G. and Keith,K.  
Plant chromosome compositions and methods  
Patent: WO 0055325-A 192 21-SEP-2000;  
The University of Chicago (US)  
FEATURES  
source  
1.47383  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
BASE COUNT 14209 a 9919 c 9688 g 13496 t 71 others  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 47383;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG.18  
|||||  
Db 7996 TGGCGCCGTTGCCAATTG 7979

RESULT 14  
AX059452/c 47840 bp DNA linear PAT 17-JAN-2001  
LOCUS  
DEFINITION  
Sequence 185 from Patent WO0055325.  
AX059452  
ACCESSION  
VERSION  
AX059452.1 GI:12311557  
KEYWORDS  
SOURCE  
ORGANISM  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE  
1 (bases 1 to 47840)

AUTHORS  
TITLE  
JOURNAL  
Preuss,D., Copenhaver,G. and Keith,K.  
Plant chromosome compositions and methods  
Patent: WO 005325-A 185 21-SEP-2000;  
The University of Chicago (US)  
FEATURES  
source  
1.47840  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
BASE COUNT 14095 a 10020 c 9567 g 14078 t 80 others  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 47840;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG.18  
|||||  
Db 35451 TGGCGCCGTTGCCAATTG 35434

RESULT 15  
AX059469 48128 bp DNA linear PAT 17-JAN-2001  
LOCUS  
DEFINITION  
Sequence 202 from Patent WO0055325.  
AX059469  
ACCESSION  
VERSION  
AX059469.1 GI:12311574  
KEYWORDS  
SOURCE  
ORGANISM  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1 (bases 1 to 48128)  
Preuss,D., Copenhaver,G. and Keith,K.  
Plant chromosome compositions and methods  
Patent: WO 005325-A 202 21-SEP-2000;  
The University of Chicago (US)  
FEATURES  
source  
1.48128  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
BASE COUNT 15779 a 7963 c 8596 g 15790 t  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 48128;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG.18  
|||||  
Db 41528 TGGCGCCGTTGCCAATTG 41545

Search completed: June 20, 2003, 23:23:56  
Job time : 77.956 secs

THIS PAGE BLANK (USPTO)



PT characters to plants, especially soybean  
 XX  
 PS Disclosure; Page 72; 118pp; English.  
 CC This oligonucleotide represents a soybean retroelement primer  
 CC binding site (version 1). The invention provides molecular tools  
 CC in the form of retroelements and retroelement-containing vectors,  
 CC cells and plants. Methods are provided for introducing the  
 CC retroelements into cells, especially when the retroelement carries  
 CC at least 1 agronomically-significant characteristic. In a  
 CC preferred method, a helper cell line which expresses gag, pol and  
 CC env sequences is used to enable transfer of a secondary construct  
 CC which carries an agronomically-significant characteristic and has  
 CC retroelement sequences that allow for replication and integration.  
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
 CC sequence selected from a retroelement primer binding site,  
 CC envelope, gag, integrase, reverse transcriptase, protease or  
 CC RNase-H sequence (see AA35254-61). Also provided are plant  
 CC retroviral particles that are used to transfer the nucleic acids  
 CC into plant cells.  
 CC  
 SQ Sequence 18 BP; 2 A; 5 C; 6 G; 5 T; 0 other;  
 XX  
 XX  
 Query Match 100.0%; Score 18; DB 21; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 8.3;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TGGCGCCGTCGCAATTG 18  
 Db 1 TGGCGCCGTCGCAATTG 18  
 XX  
 RESULT 2  
 AA35270  
 ID AA35270 standard; DNA; 149 BP.  
 XX  
 AC AA35270;  
 XX  
 DT 27-MAR-2000 (first entry)  
 XX  
 DE Plant retroelement primer binding site A.  
 XX  
 KM Retroelement; retrovirus; transgenic plant; gene transfer;  
 XX primer binding site; ss.  
 XX  
 OS Undentified.  
 XX  
 PN WO9960842-A2.  
 XX  
 PD 02-DEC-1999.  
 XX  
 PE 28-MAY-1999; 99WO-US11858.  
 XX  
 PR 29-MAY-1998; 98US-0087125.  
 XX  
 PR 28-MAY-1999; 99US-0087125.  
 XX  
 PA (WRIGHT) WRIGHT D A.  
 XX (VOYT/) VOYTAS D F.  
 XX  
 PI Wright DA, Voytas DF;  
 XX  
 DR WPI; 2000-105586/09.  
 XX  
 PT New nucleic acid molecules for imparting agronomically significant  
 PT characters to plants, especially soybean  
 XX  
 PS Disclosure; Page 115; 118pp; English.  
 CC This oligonucleotide represents plant retroelement primer binding  
 CC site A of the invention. The invention provides molecular tools  
 CC in the form of retroelements and retroelement-containing vectors,  
 CC cells and plants. Methods are provided for introducing the  
 CC retroelements into cells, especially when the retroelement carries

CC at least 1 agronomically-significant characteristic. In a  
 CC preferred method, a helper cell line which expresses gag, pol and  
 CC env sequences is used to enable transfer of a secondary construct  
 CC which carries an agronomically-significant characteristic and has  
 CC retroelement sequences that allow for replication and integration.  
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
 CC sequence selected from a retroelement primer binding site,  
 CC envelope, gag, integrase, reverse transcriptase, protease or  
 CC RNase-H sequence (see AA35254-61). Also provided are plant  
 CC retroviral particles that are used to transfer the nucleic acids  
 CC into plant cells.  
 CC  
 SQ Sequence 149 BP; 39 A; 23 C; 26 G; 61 T; 0 other;  
 XX  
 XX  
 Query Match 100.0%; Score 18; DB 21; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 9.1;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TGGCGCCGTCGCAATTG 18  
 Db 66 TGGCGCCGTCGCAATTG 83  
 XX  
 RESULT 3  
 AAF22281/C  
 ID AAF22281 standard; DNA; 59590 BP.  
 XX  
 AC AAF22281;  
 XX  
 DT 20-MAR-2001 (first entry)  
 XX  
 DE BAC containing repeats from centromeres 1-4 #4.  
 XX  
 KM Centromere; microsome; vector; ds.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO20005325-A2.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PE 17-MAR-2000; 2000WO-US07392.  
 XX  
 PR 18-MAR-1999; 99US-0125219.  
 XX  
 PR 01-APR-1999; 99US-0127409.  
 XX  
 PR 18-MAY-1999; 99US-0134770.  
 XX  
 PR 13-SEP-1999; 99US-0153584.  
 XX  
 PR 17-SEP-1999; 99US-0154603.  
 XX  
 PA (UYCH-) UNIV CHICAGO.  
 XX  
 PI Preuss D, Copenhaver G, Keith K;  
 XX  
 DR WPI; 2000-587529/55.  
 XX  
 PT Recombinant DNA construct comprising a plant centromere, useful for  
 PT producing stably inherited microsome which can serve as vectors for  
 PT the construction of transgenic plant and animal cells  
 XX  
 PS Claim 102; Page 351-364; 1449pp; English.  
 XX  
 CC The present invention relates to a recombinant DNA construct of a plant  
 CC (Arabidopsis thaliana) centromere. The constructs are useful for  
 CC producing stably inherited microsome which can serve as vectors for  
 CC the construction of transgenic plant and animal cells expressing  
 CC selected proteins such as hormones, enzymes, interleukins, clotting  
 CC factors, cytokines, antibodies, and growth factors.  
 CC  
 SQ Sequence 59590 BP; 17614 A; 12031 C; 13575 G; 16118 T; 252 other;  
 XX  
 XX  
 Query Match 100.0%; Score 18; DB 21; Length 59590;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



OY 1 TGGCGCGTGGCAATG 18  
 |||||||  
 DB 35456 TGGCGCGTGGCAATG 35439

## RESULT 4

ID AAF22279/c standard; DNA; 64415 BP.

XX AAF22279;

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #2.

XX Centromere; michromosome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells

XX Claim 102; Page 321-335; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 64415 BP; 18698 A; 13554 C; 13083 G; 18980 T; 100 other;

XX Query Match 100.0%; Score 18; DB 21; Length 64415;

XX Best Local Similarity 100.0%; Pred. No. 12;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 TGGCGCGTGGCAATG 18

XX |||||||

XX 35451 TGGCGCGTGGCAATG 35434

XX DB

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells

XX Claim 102; Page 386-404; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 57 other;

XX Query Match 100.0%; Score 18; DB 21; Length 83390;

XX Best Local Similarity 100.0%; Pred. No. 12;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 TGGCGCGTGGCAATG 18

XX |||||||

XX 71262 TGGCGCGTGGCAATG 71245

XX DB

XX RESULT 6

XX AAF22289/c standard; DNA; 90336 BP.

XX ID AAF22289 standard; DNA; 90336 BP.

XX AAF22289;

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #12.

XX Centromere; michromosome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells

XX Claim 102; Page 386-404; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 57 other;

XX Query Match 100.0%; Score 18; DB 21; Length 83390;

XX Best Local Similarity 100.0%; Pred. No. 12;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 TGGCGCGTGGCAATG 18

XX |||||||

XX 71262 TGGCGCGTGGCAATG 71245

XX DB

XX RESULT 6

XX AAF22289/c standard; DNA; 90336 BP.

XX ID AAF22289 standard; DNA; 90336 BP.

XX AAF22289;

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #12.

XX Centromere; michromosome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells

XX Claim 102; Page 386-404; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 57 other;

XX Query Match 100.0%; Score 18; DB 21; Length 83390;

XX Best Local Similarity 100.0%; Pred. No. 12;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 TGGCGCGTGGCAATG 18

XX |||||||

XX 71262 TGGCGCGTGGCAATG 71245

XX DB

XX RESULT 6

XX AAF22289/c standard; DNA; 90336 BP.

XX ID AAF22289 standard; DNA; 90336 BP.

XX AAF22289;

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #12.

XX Centromere; michromosome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells

XX Claim 102; Page 386-404; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 57 other;

XX Query Match 100.0%; Score 18; DB 21; Length 83390;

XX Best Local Similarity 100.0%; Pred. No. 12;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 TGGCGCGTGGCAATG 18

XX |||||||

XX 71262 TGGCGCGTGGCAATG 71245

XX DB

XX RESULT 6

XX AAF22289/c standard; DNA; 90336 BP.

XX ID AAF22289 standard; DNA; 90336 BP.

XX AAF22289;

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #12.

XX Centromere; michromosome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells

XX Claim 102; Page 386-404; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 57 other;

XX Query Match 100.0%; Score 18; DB 21; Length 83390;

XX Best Local Similarity 100.0%; Pred. No. 12;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 TGGCGCGTGGCAATG 18

XX |||||||

XX 71262 TGGCGCGTGGCAATG 71245

XX DB

XX RESULT 6

XX AAF22289/c standard; DNA; 90336 BP.

XX ID AAF22289 standard; DNA; 90336 BP.

XX AAF22289;

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #12.

XX Centromere; michromosome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells

XX Claim 102; Page 386

PI Preuss D, Copenhagen G, Keith K;  
XX  
XX WPI: 2000-587529/55.  
XX  
XX Recombinant DNA construct comprising a plant centromere, useful for  
PT producing stably inherited microsome which can serve as vectors for  
PT the construction of transgenic plant and animal cells  
XX  
XX  
XX  
PS Claim 102; Page 529-549; 1449pp; English.  
XX  
XX The present invention relates to a recombinant DNA construct of a plant  
CC (Arabidopsis thaliana) centromere. The constructs are useful for  
CC producing stably inherited microsome which can serve as vectors for  
CC the construction of transgenic plant and animal cells expressing  
CC selected proteins such as hormones, enzymes, interleukins, clotting  
CC factors, cytokines, antibodies, and growth factors.  
XX  
SQ Sequence 90336 BP; 26524 A; 18837 C; 18625 G; 26350 T; 0 other;  
Query Match 100.0%; Score 18; DB 21; Length 90336;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TGGCGCGCTGCCAATTG 18  
Db 84017 TGGCGCGCTGCCAATTG 84000  
RESULT 7  
AAF22288/c  
ID AAF22288 standard; DNA; 92584 BP.  
XX  
XX AAF22288;  
AC  
XX  
XX 20-MAR-2001 (first entry)  
DT  
XX  
XX BAC containing repeats from centromeres 1-4 #11.  
DE  
XX  
XX Centromere; microsome; vector; ds.  
KW  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX WO200055325-A2.  
PN  
XX  
XX 21-SEP-2000.  
PD  
XX  
XX 17-MAR-2000; 2000WO-US07392.  
PF  
XX  
XX 18-MAR-1999; 99US-0125219.  
PR  
XX  
XX 01-APR-1999; 99US-0127409.  
PR  
XX  
XX 18-MAY-1999; 99US-0134770.  
PR  
XX  
XX 13-SEP-1999; 99US-0153584.  
PR  
XX  
XX 17-SEP-1999; 99US-0154603.  
XX  
XX  
XX (UYCH-) UNIV CHICAGO.  
PA  
XX  
XX Preuss D, Copenhagen G, Keith K;  
PI  
XX  
XX WPI: 2000-587529/55.  
DR  
XX  
XX Recombinant DNA construct comprising a plant centromere, useful for  
PT producing stably inherited microsome which can serve as vectors for  
PT the construction of transgenic plant and animal cells  
XX  
XX  
XX  
PS Claim 102; Page 508-529; 1449pp; English.  
XX  
XX The present invention relates to a recombinant DNA construct of a plant  
CC (Arabidopsis thaliana) centromere. The constructs are useful for  
CC producing stably inherited microsome which can serve as vectors for  
CC the construction of transgenic plant and animal cells expressing  
CC selected proteins such as hormones, enzymes, interleukins, clotting  
CC factors, cytokines, antibodies, and growth factors.  
XX

SQ Sequence 92584 BP; 27840 A; 18113 C; 19835 G; 26796 T; 0 other;  
Query Match 100.0%; Score 18; DB 21; Length 92584;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TGGCGCGCTGCCAATTG 18  
Db 31015 TGGCGCGCTGCCAATTG 30998  
RESULT 8  
AAF22302/c  
ID AAF22302 standard; DNA; 94895 BP.  
XX  
XX AAF22302;  
AC  
XX  
XX 20-MAR-2001 (first entry)  
DT  
XX  
XX BAC containing repeats from centromeres 1-4 #25.  
DE  
XX  
XX Centromere; microsome; vector; ds.  
KW  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX WO200055325-A2.  
PN  
XX  
XX 21-SEP-2000.  
PD  
XX  
XX 17-MAR-2000; 2000WO-US07392.  
PF  
XX  
XX 18-MAR-1999; 99US-0125219.  
PR  
XX  
XX 01-APR-1999; 99US-0127409.  
PR  
XX  
XX 18-MAY-1999; 99US-0134770.  
PR  
XX  
XX 13-SEP-1999; 99US-0153584.  
PR  
XX  
XX 17-SEP-1999; 99US-0154603.  
XX  
XX  
XX (UYCH-) UNIV CHICAGO.  
PA  
XX  
XX Preuss D, Copenhagen G, Keith K;  
PI  
XX  
XX WPI: 2000-587529/55.  
DR  
XX  
XX Recombinant DNA construct comprising a plant centromere, useful for  
PT producing stably inherited microsome which can serve as vectors for  
PT the construction of transgenic plant and animal cells  
XX  
XX  
XX  
PS Claim 102; Page 804-819; 1449pp; English.  
XX  
XX The present invention relates to a recombinant DNA construct of a plant  
CC (Arabidopsis thaliana) centromere. The constructs are useful for  
CC producing stably inherited microsome which can serve as vectors for  
CC the construction of transgenic plant and animal cells expressing  
CC selected proteins such as hormones, enzymes, interleukins, clotting  
CC factors, cytokines, antibodies, and growth factors.  
XX  
XX  
XX  
SQ Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;  
Query Match 100.0%; Score 18; DB 21; Length 94895;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TGGCGCGCTGCCAATTG 18  
Db 38282 TGGCGCGCTGCCAATTG 38265  
RESULT 9  
AAF22282/c  
ID AAF22282 standard; DNA; 95223 BP.  
XX  
XX AAF22282;  
AC  
XX

```

DT 20-MAR-2001 (first entry)
XX BAC containing repeats from centromeres 1-4 #5.
DE Centromere; microsome; vector; ds.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX MO200055325-A2.
XX 21-SEP-2000.
XX 17-MAR-2000; 2000WO-US07392.
XX 18-MAR-1999; 99US-0125219.
XX 01-APR-1999; 99US-0127409.
XX 18-MAY-1999; 99US-0134770.
XX 13-SEP-1999; 99US-0153584.
XX 17-SEP-1999; 99US-0154603.
XX (UYCH-) UNIV CHICAGO.
XX Preuss D, Copenhaver G, Keith K;
XX WPI; 2000-587529/55.
XX Recombinant DNA construct comprising a plant centromere, useful for
XX producing stably inherited microsome which can serve as vectors for
XX the construction of transgenic plant and animal cells
XX
XX Claim 102; Page 364-385; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 95223 BP; 27974 A; 19452 C; 19355 G; 27745 T; 697 other;
Query Match 100.0%; Score 18; DB 21; Length 95223;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGGCGCCGTTGCCAATTG 18
Db 55836 TGGCGCCGTTGCCAATTG 55819

```

```

PR 13-SEP-1999; 99US-0153584.
XX 17-SEP-1999; 99US-0154603.
XX (UYCH-) UNIV CHICAGO.
XX Preuss D, Copenhaver G, Keith K;
XX WPI; 2000-587529/55.
XX Recombinant DNA construct comprising a plant centromere, useful for
XX producing stably inherited microsome which can serve as vectors for
XX the construction of transgenic plant and animal cells
XX
XX Claim 102; Page 716-738; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 96583 BP; 29910 A; 18944 C; 18341 G; 29388 T; 0 other;
Query Match 100.0%; Score 18; DB 21; Length 96583;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGGCGCCGTTGCCAATTG 18
Db 1458 TGGCGCCGTTGCCAATTG 1441

```

```

RESULT 10
AAF22297/C
ID AAF22297 standard; DNA; 96583 BP.
XX AAF22297;
AC 20-MAR-2001 (first entry)
XX 20-MAR-2001 (first entry)
DE BAC containing repeats from centromeres 1-4 #20.
XX Centromere; microsome; vector; ds.
XX Arabidopsis thaliana.
XX Arabidopsis thaliana.
XX MO200055325-A2.
XX 21-SEP-2000.
XX 17-MAR-2000; 2000WO-US07392.
XX 18-MAR-1999; 99US-0125219.
XX 01-APR-1999; 99US-0127409.
XX 18-MAY-1999; 99US-0134770.

```

```

RESULT 11
AAF22290
ID AAF22290 standard; DNA; 96988 BP.
XX AAF22290;
AC 20-MAR-2001 (first entry)
XX 20-MAR-2001 (first entry)
DE BAC containing repeats from centromeres 1-4 #13.
XX Centromere; microsome; vector; ds.
XX Arabidopsis thaliana.
XX Arabidopsis thaliana.
XX MO200055325-A2.
XX 21-SEP-2000.
XX 17-MAR-2000; 2000WO-US07392.
XX 18-MAR-1999; 99US-0125219.
XX 01-APR-1999; 99US-0127409.
XX 18-MAY-1999; 99US-0134770.
XX 13-SEP-1999; 99US-0153584.
XX 17-SEP-1999; 99US-0154603.
XX (UYCH-) UNIV CHICAGO.
XX Preuss D, Copenhaver G, Keith K;
XX WPI; 2000-587529/55.
XX Recombinant DNA construct comprising a plant centromere, useful for
XX producing stably inherited microsome which can serve as vectors for
XX the construction of transgenic plant and animal cells
XX
XX Claim 102; Page 550-572; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for

```

CC producing stably inherited mitochondria which can serve as vectors for  
CC the construction of transgenic plant and animal cells expressing  
CC selected proteins such as hormones, enzymes, interleukins, clotting  
CC factors, cytokines, antibodies, and growth factors.

Sequence 96988 BP; 29654 A; 18289 C; 18523 G; 30522 T; 0 other;

Query Match 100.0%; Score 18; DB 21; Length 96988;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18  
DB 49874 TGGCGCCGTTGCCAATTG 49891

RESULT 12

AAAF22298 standard; DNA; 109973 BP.

AC AAAF22298;

DT 20-MAR-2001 (first entry)

DE BAC containing repeats from centromeres 1-4 #21.

KW Centromere; mitochondria; vector; ds.

OS Arabidopsis thaliana.

PN WO200055325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000MO-US07392.

PR 18-MAR-1999; 99US-0125219.

PR 01-APR-1999; 99US-0127409.

PR 18-MAY-1999; 99US-0134770.

PR 13-SEP-1999; 99US-0153584.

PR 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

PI Preuss D, Copenhaver G, Keith K;

DR WPI; 2000-587529/55.

PT Recombinant DNA construct comprising a plant centromere, useful for

PT producing stably inherited mitochondria which can serve as vectors for

PT the construction of transgenic plant and animal cells

PS Claim 102; Page 738-763; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

CC (Arabidopsis thaliana) centromere. The constructs are useful for

CC producing stably inherited mitochondria which can serve as vectors for

CC the construction of transgenic plant and animal cells expressing

CC selected proteins such as hormones, enzymes, interleukins, clotting

CC factors, cytokines, antibodies, and growth factors.

XX Sequence 109973 BP; 31657 A; 23911 C; 22655 G; 31745 T; 5 other;

Query Match 100.0%; Score 18; DB 21; Length 109973;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AAAF22286 standard; DNA; 134499 BP.

AC AAAF22286;

DT 20-MAR-2001 (first entry)

DE BAC containing repeats from centromeres 1-4 #9.

KW Centromere; mitochondria; vector; ds.

OS Arabidopsis thaliana.

PN WO200055325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000MO-US07392.

PR 18-MAR-1999; 99US-0125219.

PR 01-APR-1999; 99US-0127409.

PR 18-MAY-1999; 99US-0134770.

PR 13-SEP-1999; 99US-0153584.

PR 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

PI Preuss D, Copenhaver G, Keith K;

DR WPI; 2000-587529/55.

PT Recombinant DNA construct comprising a plant centromere, useful for

PT producing stably inherited mitochondria which can serve as vectors for

PT the construction of transgenic plant and animal cells

PS Claim 102; Page 453-484; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

CC (Arabidopsis thaliana) centromere. The constructs are useful for

CC producing stably inherited mitochondria which can serve as vectors for

CC the construction of transgenic plant and animal cells expressing

CC selected proteins such as hormones, enzymes, interleukins, clotting

CC factors, cytokines, antibodies, and growth factors.

XX Sequence 134499 BP; 41565 A; 25130 C; 25225 G; 42577 T; 2 other;

Query Match 100.0%; Score 18; DB 21; Length 134499;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18

DB 89656 TGGCGCCGTTGCCAATTG 89673

RESULT 14

AAAF22303/C

ID AAAF22303 standard; DNA; 611590 BP.

AC AAAF22303;

DT 20-MAR-2001 (first entry)

DE Arabidopsis thaliana chromosome 2 centromere.

KW Centromere; mitochondria; vector; ds.

OS Arabidopsis thaliana.

PN WO200055325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000WO-US07392.  
XX  
PR 18-MAR-1999; 99US-0125219.  
PR 01-APR-1999; 99US-0127409.  
PR 18-MAY-1999; 99US-0134770.  
PR 13-SEP-1999; 99US-0153584.  
PR 17-SEP-1999; 99US-0154603.  
XX  
PA (UYCH-) UNIV CHICAGO.  
XX  
PI Preuss D, Copenhaver G, Keith K;  
XX  
DR WPI; 2000-587529/55.  
XX  
PT Recombinant DNA construct comprising a plant centromere, useful for  
PT producing stably inherited microchromosomes which can serve as vectors for  
PT the construction of transgenic plant and animal cells  
XX  
PS Claim 45; Page 820-959; 1449pp; English.  
XX  
CC The present invention relates to a recombinant DNA construct of a plant  
CC (Arabidopsis thaliana) centromere. The constructs are useful for  
CC producing stably inherited microchromosomes which can serve as vectors for  
CC the construction of transgenic plant and animal cells expressing  
CC selected proteins such as hormones, enzymes, interleukins, clotting  
CC factors, cytokines, antibodies, and growth factors.  
XX  
SQ Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 184983 T; 0 other;  
OY  
Query Match 100.0%; Score 18; DB 21; Length 611590;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 TGGCGCCGTTGCCAATTG 18  
|||||  
94448 TGGCGCCGTTGCCAATTG 94431  
DB  
RESULT 15  
AAF22305  
ID AAF22305 standard; DNA; 1082138 BP.  
XX  
AC AAF22305;  
XX  
DT 20-MAR-2001 (first entry)  
XX  
DE Arabidopsis thaliana chromosome 4 centromere.  
XX  
KW Centromere; microsome; vector; ds.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO200055325-A2.  
XX  
PD 21-SEP-2000.  
XX  
PE 17-MAR-2000; 2000WO-US07392.  
XX  
PR 18-MAR-1999; 99US-0125219.  
PR 01-APR-1999; 99US-0127409.  
PR 18-MAY-1999; 99US-0134770.  
PR 13-SEP-1999; 99US-0153584.  
PR 17-SEP-1999; 99US-0154603.  
XX  
PA (UYCH-) UNIV CHICAGO.  
XX  
PI Preuss D, Copenhaver G, Keith K;  
XX  
DR WPI; 2000-587529/55.  
XX  
PT Recombinant DNA construct comprising a plant centromere, useful for  
PT producing stably inherited microchromosomes which can serve as vectors for  
PT the construction of transgenic plant and animal cells

XX  
PS Claim 68; Page 977-1388; 1449pp; English.  
XX  
CC The present invention relates to a recombinant DNA construct of a plant  
CC (Arabidopsis thaliana) centromere. The constructs are useful for  
CC producing stably inherited microchromosomes which can serve as vectors for  
CC the construction of transgenic plant and animal cells expressing  
CC selected proteins such as hormones, enzymes, interleukins, clotting  
CC factors, cytokines, antibodies, and growth factors.  
XX  
SQ Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;  
OY  
Query Match 100.0%; Score 18; DB 21; Length 1082138;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 TGGCGCCGTTGCCAATTG 18  
|||||  
1043856 TGGCGCCGTTGCCAATTG 1043873  
DB

Search completed: June 20, 2003, 21:58:19  
Job time : 13.3019 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:44:01 ; Search time 2.17925 Seconds  
(without alignments)  
2533.070 Million cell updates/sec

Title: US-09-965-553-1  
Perfect score: 18  
Sequence: 1 tggcgcgcgttgcattg 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: Issued\_Patents\_NA:\*  
2: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/Backfiles1.seq:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	18	US-09-322-478-1	Sequence 1, Appl
2	18	100.0	150	US-09-322-478-36	Sequence 36, Appl
3	15.4	85.6	1713	US-08-386-727-5	Sequence 5, Appl
4	15.4	85.6	1713	US-08-600-452A-5	Sequence 5, Appl
5	14.8	82.2	1272	US-09-053-702-1	Sequence 1, Appl
6	14.8	82.2	3097	US-09-282-147-38	Sequence 38, Appl
7	14.4	80.0	962	US-09-072-596-310	Sequence 310, Appl
8	14.4	80.0	9515	US-08-920-812-13	Sequence 13, Appl
9	14.4	80.0	9515	US-08-920-827-13	Sequence 13, Appl
10	14.4	80.0	9515	US-08-921-177-13	Sequence 13, Appl
11	14.4	80.0	9515	US-08-362-577C-13	Sequence 13, Appl
12	14.4	80.0	9515	US-08-920-828-13	Sequence 13, Appl
13	14.4	80.0	4403765	US-09-103-840A-2	Sequence 2, Appl
14	14.4	80.0	4403765	US-09-103-840A-2	Sequence 2, Appl
15	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
16	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
17	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
18	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
19	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
20	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
21	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
22	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
23	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
24	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
25	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
26	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl
27	14.4	80.0	4411529	US-09-103-840A-1	Sequence 1, Appl

28	13.8	76.7	1659	US-08-548-509-4	Sequence 4, Appl
29	13.8	76.7	3538	US-09-336-447A-10	Sequence 10, Appl
30	13.8	76.7	4588	US-09-453-702B-26	Sequence 26, Appl
31	13.8	76.7	10684	US-09-221-017B-401	Sequence 401, Appl
32	13.8	76.7	24417	US-08-846-762-1	Sequence 1, Appl
33	13.4	74.4	1596	US-08-531-601-2	Sequence 2, Appl
34	13.4	74.4	1596	US-08-859-032-2	Sequence 2, Appl
35	13.4	74.4	3390	US-09-550-338-1	Sequence 1, Appl
36	13.4	74.4	3825	US-08-737-597-1	Sequence 1, Appl
37	13.4	74.4	3825	US-08-737-597-2	Sequence 2, Appl
38	13.4	74.4	6312	US-08-531-601-3	Sequence 3, Appl
39	13.4	74.4	6312	US-08-859-032-3	Sequence 3, Appl
40	13.4	74.4	68750	US-09-335-409-1	Sequence 1, Appl
41	13.4	74.4	68750	US-09-568-102-1	Sequence 1, Appl
42	13.4	74.4	68750	US-09-567-969-1	Sequence 1, Appl
43	13.4	74.4	68750	US-09-568-480-1	Sequence 1, Appl
44	13.4	74.4	68750	US-09-568-486-1	Sequence 1, Appl
45	13.4	74.4	68750	US-09-568-472-1	Sequence 1, Appl

## ALIGNMENTS

```

RESULT 1
US-09-322-478-1
; Sequence 1, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; EARLIER FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-1

Query Match      100.0%; Score 18; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. NO. 0.74;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCGGTGCCAATG 18
DB      1 TGGCGCGGTGCCAATG 18

RESULT 2
US-09-322-478-36
; Sequence 36, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; EARLIER FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-36

```

Query Match 100.0%; Score 18; DB 4; Length 150;  
Best Local Similarity 100.0%; Pred. No. 0.94;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCGTGGCAATTG 18  
|||||  
Db 66 TGGCGCGTGGCAATTG 83

## RESULT 3

US-08-386-727-5/c

Sequence 5, Application US/08386727  
Patent No. 5792647

GENERAL INFORMATION:

APPLICANT: ROSEMAN, SAUL

APPLICANT: BASSLER, BONNIE

APPLICANT: KEYHANT, NEMAT O.

APPLICANT: CHITLARD, EDITH

APPLICANT: ROME, CHRIS

APPLICANT: YU, CHARLES

TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

ADDRESSEE: CUSHMAN, DARBY &amp; CUSHMAN

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/386,727

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: HOBBS, ANN S.

REGISTRATION NUMBER: 36,830

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-861-3000

TELEFAX: 202-822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1713 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-386-727-5

QY 2 GGGCGCGTGGCAATTG 18  
|||||  
Db 591 GGGCGCGTGGCAATTG 575

## RESULT 4

US-08-600-452A-5/c

Sequence 5, Application US/08600452A  
Patent No. 5985644

GENERAL INFORMATION:

APPLICANT: ROSEMAN, SAUL

APPLICANT: BASSLER, BONNIE

APPLICANT: KEYHANT, NEMAT O.

APPLICANT: CHITLARD, EDITH  
APPLICANT: ROME, CHRIS  
APPLICANT: YU, CHARLES  
TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:

ADDRESSEE: FISH &amp; RICHARDSON P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/600,452A

FILING DATE: 13-FEB-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hallie, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 07662/005001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 678-5070

TELEFAX: (619) 678-5099

TELEX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1713 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-600-452A-5

Query Match 85.6%; Score 15.4; DB 2; Length 1713;  
Best Local Similarity 94.1%; Pred. No. 28;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGCGCGTGGCAATTG 18  
|||||  
Db 591 GGGCGCGTGGCAATTG 575

## RESULT 5

US-09-053-702-1

Sequence 1, Application US/09053702  
Patent No. 6229069

GENERAL INFORMATION:

APPLICANT: YAMADA, Shigehiro

TITLE OF INVENTION: METHOD FOR CONTROLLING WATER CONTENT OF PLANT

FILE REFERENCE: 230-122P

CURRENT APPLICATION NUMBER: US/09/053,702

CURRENT FILING DATE: 1998-04-02

NUMBER OF SEQ ID NOS: 5

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO: 1

LENGTH: 1272

TYPE: DNA

ORGANISM: Mesembryanthemum crystallinum

FEATURE:

NAME/KEY: CDS

LOCATION: (225)..(1070)

US-09-053-702-1

Query Match 82.2%; Score 14.8; DB 4; Length 1272;  
Best Local Similarity 88.9%; Pred. No. 56;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCGTGGCAATTG 18



DB 841 TGGCTCCATTGCCAATTG 858

## RESULT 6

US-09-282-147-38/C  
Sequence 38, Application US/09282147  
Patent No. 6274147  
GENERAL INFORMATION:  
APPLICANT: VAKHARIA, VIKRAM  
APPLICANT: YAO, KUN  
TITLE OF INVENTION: METHOD FOR GENERATING NONPATHOGENIC, INFECTIOUS  
TITLE OF INVENTION: PANCREATIC NECROSIS VIRUS (IPNV) FROM SYNTHETIC RNA  
FILE REFERENCE: 8288-9023  
CURRENT APPLICATION NUMBER: US/09/282,147  
CURRENT FILING DATE: 1999-03-31  
EARLIER APPLICATION NUMBER: US/60/080,278  
EARLIER FILING DATE: 1998-03-31  
EARLIER APPLICATION NUMBER: PCT/US97/12955  
EARLIER FILING DATE: 1998-03-31  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 38  
LENGTH: 3097  
TYPE: DNA  
ORGANISM: Infectious pancreatic necrosis virus  
US-09-282-147-38

Query Match 82.2%; Score 14.8; DB 4; Length 3097;  
Best Local Similarity 88.9%; Pred. No. 62;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18  
DB 1091 TGGCGACGTTGCCGATTG 1074

## RESULT 7

US-09-072-596-310/C  
Sequence 310, Application US/09072596  
Patent No. 6458366  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonia  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
NUMBER OF SEQUENCES: 350  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,596  
FILING DATE: 05-MAY-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 310:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 962 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-09-072-596-310

Query Match 80.0%; Score 14.4; DB 4; Length 962;  
Best Local Similarity 93.8%; Pred. No. 88;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAAT 16  
DB 142 TGGCGCCGTTGCCGAT 127

## RESULT 8

US-08-920-812-13/C  
Sequence 13, Application US/08920812  
Patent No. 5763188  
GENERAL INFORMATION:-

APPLICANT: Ohno, Tsuneya  
APPLICANT: Matsunisa, Akio  
APPLICANT: Uehara, Hirotsugu  
APPLICANT: Eda, Soji  
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/920,812  
FILING DATE: 29-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,577  
FILING DATE: 27-MAR-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 19036/32420  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9515 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas aeruginosa  
STRAIN: Clinical isolate P2-2  
US-08-920-812-13

Query Match 80.0%; Score 14.4; DB 1; Length 9515;

Best Local Similarity 93.8%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCGCTTGCCAAATG 18  
|||||  
Db 515 GCGCGCTTGCCAAATCG 500

## RESULT 9

US-08-920-827-13/C  
; Sequence 13, Application US/08920827  
; Patent No. 5770375  
; GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya  
APPLICANT: Matsuhisa, Akio  
APPLICANT: Uehara, Hirotsugu

APPLICANT: Eda, Soji  
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago

STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/920,827

FILING DATE: 29-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,577  
FILING DATE: 27-MAR-1995  
ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 19036/32420

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448

TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:

LENGTH: 9515 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:

ORGANISM: Pseudomonas aeruginosa  
STRAIN: Clinical Isolate P2-2  
US-08-920-827-13

Query Match 80.0%; Score 14.4; DB 1; Length 9515;  
Best Local Similarity 93.8%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCGCTTGCCAAATG 18  
|||||  
Db 515 GCGCGCTTGCCAAATCG 500

RESULT 10  
US-08-921-177-13/C  
; Sequence 13, Application US/08921177  
; Patent No. 5798211  
; GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya

APPLICANT: Matsuhisa, Akio

APPLICANT: Uehara, Hirotsugu

APPLICANT: Matsuhisa, Akio  
APPLICANT: Uehara, Hirotsugu  
APPLICANT: Eda, Soji  
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago

STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/921,177

FILING DATE: 29-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,577  
FILING DATE: 27-MAR-1995  
ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 19036/32420

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448

TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:

LENGTH: 9515 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:

ORGANISM: Pseudomonas aeruginosa  
STRAIN: Clinical Isolate P2-2  
US-08-921-177-13

Query Match 80.0%; Score 14.4; DB 1; Length 9515;  
Best Local Similarity 93.8%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCGCTTGCCAAATG 18  
|||||  
Db 515 GCGCGCTTGCCAAATCG 500

RESULT 11  
US-08-362-577C-13/C  
; Sequence 13, Application US/08362577C  
; Patent No. 5807673  
; GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya  
APPLICANT: Matsuhisa, Akio  
APPLICANT: Uehara, Hirotsugu

APPLICANT: Eda, Soji  
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago

STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,577C  
FILING DATE: 27-MAR-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 19036/32420  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ. ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9515 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas aeruginosa  
STRAIN: Clinical Isolate P2-2  
US-08-362-577C-13

Query Match 80.0%; Score 14.4; DB 1; Length 9515;  
Best Local Similarity 93.8%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCGCTTGCCATTG 18  
|||||  
Db 515 GCGCGCTTGCCATTG 500

RESULT 12  
US-08-920-828-13/C  
Sequence 13, Application US/08920828  
Patent No. 5853998  
GENERAL INFORMATION:  
APPLICANT: Ohno, Tsuneya  
APPLICANT: Matsuhisa, Akio  
APPLICANT: Uehara, Hirotsugu  
APPLICANT: Edo, Soji  
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/920,828  
FILING DATE: 29-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,577  
FILING DATE: 27-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 19036/32420  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448

TELEX: 25-3856  
INFORMATION FOR SEQ. ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9515 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas aeruginosa  
STRAIN: Clinical Isolate P2-2  
US-08-920-828-13

Query Match 80.0%; Score 14.4; DB 2; Length 9515;  
Best Local Similarity 93.8%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCGCTTGCCATTG 18  
|||||  
Db 515 GCGCGCTTGCCATTG 500

RESULT 13  
US-09-103-840A-2  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
US-09-103-840A-2

Query Match 80.0%; Score 14.4; DB 4; Length 4403765;  
Best Local Similarity 93.8%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCGCGCTTGCCAAAT 16  
|||||  
Db 1572527 TGCGCGCTTGCCAAAT 1572542

RESULT 14  
US-09-103-840A-2/C  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 80.0%; Score 14.4; DB 4; Length 4403765;  
Best Local Similarity 93.8%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGCCAAT 16  
|||  
Db 841864 TGGCGCGGTGCCAAT 841849

RESULT 15  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1

Query Match 80.0%; Score 14.4; DB 4; Length 4411529;  
Best Local Similarity 93.8%; Pred. No. 67;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGCCAAT 16  
|||  
Db 1572688 TGGCGCGGTGCCAAT 1572703

Search completed: June 20, 2003, 23:25:32  
Job time: 12.1792 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:58:41 ; Search time 6.65094 Seconds  
(without alignments)  
3971.420 Million cell updates/sec

Title: US-09-965-553-1

Perfect score: 18

Sequence: 1 tggcgccgttgccattg 18

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.\*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	18	10	US-09-965-553-1
2	18	100.0	150	10	US-09-965-553-36
3	15	83.3	412	9	US-10-101-464A-479
4	14.8	82.2	266	10	US-09-878-574-15377
5	14.8	82.2	573	10	US-09-974-300-4705
6	14.8	82.2	861	9	US-09-938-842A-807
7	14.8	82.2	1089	10	US-09-974-300-764
8	14.8	82.2	2155	9	US-10-002-050-11
9	14.8	82.2	2155	9	US-10-002-304-11
10	14.8	82.2	2155	12	US-10-003-152-11
11	14.8	82.2	2156	9	US-10-002-050-21
12	14.8	82.2	2156	9	US-10-002-304-21
13	14.8	82.2	2156	12	US-10-003-152-21
14	14.8	82.2	2284	9	US-10-002-050-13
15	14.8	82.2	2284	9	US-10-002-304-13
16	14.8	82.2	2284	12	US-10-003-152-13
17	14.8	82.2	3097	9	US-10-216-981A-25
18	14.8	82.2	3293	9	US-10-149-819-25
19	14.8	82.2	8095	9	US-09-989-920-73

C	20	14.4	80.0	943	10	US-09-822-830A-533	Sequence 533, App
C	21	14.4	80.0	1854	9	US-09-894-844-103	Sequence 103, App
C	22	14.4	80.0	1863	10	US-09-815-242-9885	Sequence 9885, App
C	23	14	77.8	1392	10	US-09-815-242-9820	Sequence 9820, App
C	24	14	77.8	3268	9	US-09-927-827-24	Sequence 24, App1
C	25	13.8	76.7	174	9	US-10-122-822-21	Sequence 21, App1
C	26	13.8	76.7	365	10	US-09-880-107-357	Sequence 357, App
C	27	13.8	76.7	476	10	US-09-974-300-5603	Sequence 5603, App
C	28	13.8	76.7	538	9	US-09-925-299-369	Sequence 369, App
C	29	13.8	76.7	538	10	US-09-925-299-369	Sequence 369, App
C	30	13.8	76.7	576	10	US-09-974-300-5436	Sequence 5436, App
C	31	13.8	76.7	897	10	US-09-841-132-120	Sequence 120, App
C	32	13.8	76.7	897	10	US-09-841-132-130	Sequence 130, App
C	33	13.8	76.7	897	10	US-09-841-132-132	Sequence 132, App
C	34	13.8	76.7	1078	10	US-09-974-300-2700	Sequence 2700, App
C	35	13.8	76.7	1320	10	US-09-815-242-9819	Sequence 9819, App
C	36	13.8	76.7	1405	9	US-10-122-822-6	Sequence 6, App1
C	37	13.8	76.7	1627	12	US-10-062-254-237	Sequence 237, App
C	38	13.8	76.7	2940	10	US-09-801-368-283	Sequence 283, App
C	39	13.8	76.7	3538	9	US-09-952-267-10	Sequence 10, App1
C	40	13.8	76.7	4588	9	US-10-114-170-26	Sequence 26, App1
C	41	13.8	76.7	6373	9	US-10-000-512-1	Sequence 1, App1
C	42	13.8	76.7	6378	9	US-10-037-270-332	Sequence 332, App
C	43	13.8	76.7	13715	7	US-08-781-986A-195	Sequence 195, App
C	44	13.8	76.7	465237	10	US-09-933-267A-1	Sequence 1, App1
C	45	13.4	74.4	277	10	US-09-294-093B-2256	Sequence 2256, App

#### ALIGNMENTS

```
RESULT 1
US-09-965-553-1
; Sequence 1, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Voytas, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-1

Query Match      100.0%  Score 18; DB 10; Length 18;
Best Local Similarity 100.0%  Pred No. 1.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGCCAATTG 18
      |||||||
DB      1 TGGCGCCGTTGCCAATTG 18

RESULT 2
US-09-965-553-36
; Sequence 36, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
```

;; CURRENT FILING DATE: 2001-09-27  
;; PRIOR APPLICATION NUMBER: 09/322,478  
;; PRIOR FILING DATE: 1999-05-28  
;; PRIOR APPLICATION NUMBER: 60/087125  
;; PRIOR FILING DATE: 1998-05-29  
;; NUMBER OF SEQ ID NOS: 41  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 36  
;; LENGTH: 150  
;; TYPE: DNA  
;; ORGANISM: Glycine max  
US-09-965-553-36

Query Match 100.0%; Score 18; DB 10; Length 150;  
Best Local Similarity 100.0%; Pred. NO. 2.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGCGCGTTGCCAATTG 18  
|||||  
DB 66 TGGGCGCGTTGCCAATTG 83

RESULT 3  
US-10-101-464A-479/C  
;; Sequence 479, Application US/10101464A  
;; Publication No. US20030046728A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Strabala, Timothy  
;; APPLICANT: Nieuwenhuizen, Nicolaas  
;; APPLICANT: Higgins, Colleen M.  
;; TITLE OF INVENTION: Compositions isolated from plant cells  
;; TITLE OF INVENTION: and their use in the modification of plant cell signaling  
;; FILE REFERENCE: 11000.1020c2  
;; CURRENT APPLICATION NUMBER: US/10/101,464A  
;; CURRENT FILING DATE: 2002-03-18  
;; PRIOR APPLICATION NUMBER: 09/704,302  
;; PRIOR FILING DATE: 2000-11-01  
;; PRIOR APPLICATION NUMBER: 09/228,986  
;; PRIOR FILING DATE: 1999-01-12  
;; PRIOR APPLICATION NUMBER: 60/162,866  
;; PRIOR FILING DATE: 1999-11-01  
;; PRIOR APPLICATION NUMBER: PCT/US00/00724  
;; PRIOR FILING DATE: 2000-01-11  
;; NUMBER OF SEQ ID NOS: 989  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 479  
;; LENGTH: 412  
;; TYPE: DNA  
;; ORGANISM: Pinus radiata  
US-10-101-464A-479

Query Match 83.3%; Score 15; DB 9; Length 412;  
Best Local Similarity 100.0%; Pred. NO. 1e+02;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCGCGCTTGCCATT 17  
|||||  
DB 32 GCGCGCTTGCCATT 18

RESULT 4  
US-09-878-574-15377  
;; Sequence 15377, Application US/09878574  
;; Patent No. US20020110548A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Byrum, Joseph R.  
;; APPLICANT: La Rosa, Thomas J.  
;; APPLICANT: Thompson, Michael D.  
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
;; TITLE OF INVENTION: Plants  
;; FILE REFERENCE: 38-21115401B  
;; CURRENT APPLICATION NUMBER: US/09/878,574  
;; CURRENT FILING DATE: 2001-12-21

;; PRIOR APPLICATION NUMBER: 09/333,535  
;; PRIOR FILING DATE: 1999-06-14  
;; NUMBER OF SEQ ID NOS: 15775  
;; SEQ ID NO 15377  
;; LENGTH: 266  
;; TYPE: DNA  
;; ORGANISM: Glycine max  
;; OTHER INFORMATION: Clone ID: 701070032H2  
US-09-878-574-15377

Query Match 82.2%; Score 14.8; DB 10; Length 266;  
Best Local Similarity 88.9%; Pred. NO. 1.3e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGGCGCGTTGCCAATTG 18  
|||||  
DB 6 TGGCACCCTTGCCATTG 23

RESULT 5  
US-09-974-300-4705  
;; Sequence 4705, Application US/09974300  
;; Patent No. US20020146721A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Berka, Randy M.  
;; APPLICANT: Clausen, Ib Groth  
;; TITLE OF INVENTION: Methods for Monitoring Multiple Gene  
;; TITLE OF INVENTION: Expression  
;; FILE REFERENCE: 10085.500-US  
;; CURRENT APPLICATION NUMBER: US/09/974,300  
;; CURRENT FILING DATE: 2001-10-05  
;; PRIOR APPLICATION NUMBER: 09/680,598  
;; PRIOR FILING DATE: 2000-10-06  
;; PRIOR APPLICATION NUMBER: 60/279,526  
;; PRIOR FILING DATE: 2001-03-27  
;; NUMBER OF SEQ ID NOS: 8481  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4705  
;; LENGTH: 573  
;; TYPE: DNA  
;; ORGANISM: Bacillus clausii  
US-09-974-300-4705

Query Match 82.2%; Score 14.8; DB 10; Length 573;  
Best Local Similarity 88.9%; Pred. NO. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGGCGCGTTGCCAATTG 18  
|||||  
DB 287 TGACGCGCTTGCGATTG 304

RESULT 6  
US-09-938-842A-807  
;; Sequence 807, Application US/09938842A  
;; Patent No. US20020160378A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Harper, Jeff  
;; APPLICANT: Kreps, Joel  
;; APPLICANT: Wang, Xun  
;; APPLICANT: Zhu, Tong  
;; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
;; TITLE OF INVENTION: SAME, AND METHODS OF USE  
;; FILE REFERENCE: SCRIPL300-3  
;; CURRENT APPLICATION NUMBER: US/09/938,842A  
;; CURRENT FILING DATE: 2001-08-24  
;; PRIOR APPLICATION NUMBER: US 60/227,866  
;; PRIOR FILING DATE: 2000-08-24  
;; PRIOR APPLICATION NUMBER: US 60/264,647  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: US 60/300,111  
;; PRIOR FILING DATE: 2001-06-22  
;; NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 807  
LENGTH: 861  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-807

Query Match  
Best Local Similarity 82.2%; Score 14.8; DB 9; Length 861;  
Best Local Similarity 88.9%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
DB 605 TGGCTCCATGCGCAATTG 622

RESULT 7  
US-09-974-300-764/c  
Sequence 764, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
APPLICANT: Clausen, Id Groth  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
FILE REFERENCE: 10085, 500-US  
CURRENT APPLICATION NUMBER: US/09/974, 300  
PRIOR FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680, 598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279, 526  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 764  
LENGTH: 1089  
TYPE: DNA  
ORGANISM: Bacillus licheniformis  
US-09-974-300-764

Query Match  
Best Local Similarity 82.2%; Score 14.8; DB 10; Length 1089;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
DB 1070 TGGCGCCGTTGCCAATTG 1053

RESULT 8  
US-10-002-050-11/c  
Sequence 11, Application US/10002050  
Publication No. US20030032095A1  
GENERAL INFORMATION:  
APPLICANT: Shimkets, Richard  
APPLICANT: Fernandes, Elma  
APPLICANT: Vernet, Corine  
APPLICANT: Yang, Meijia  
APPLICANT: Boldog, Ferenc  
APPLICANT: Herrmann, John  
TITLE OF INVENTION: No. US20030032095A1 Nucleic Acid Sequences Encoding Human Semaphorin  
FILE REFERENCE: 15966-554 Cura-54 CON-S14  
CURRENT APPLICATION NUMBER: US/10/002, 050  
PRIOR FILING DATE: 2001-11-02  
PRIOR APPLICATION NUMBER: 09/604, 286  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 60/140, 584  
PRIOR FILING DATE: 1999-06-23  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 2155  
TYPE: DNA  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: CDS  
LOCATION: (166)..(1935)  
US-10-002-050-11

Query Match  
Best Local Similarity 82.2%; Score 14.8; DB 9; Length 2155;  
Best Local Similarity 88.9%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
DB 1298 TGGCGCCGATGCCAGTTG 1281

RESULT 9  
US-10-002-304-11/c  
Sequence 11, Application US/10002304  
Publication No. US20030036185A1  
GENERAL INFORMATION:  
APPLICANT: Shimkets, Richard  
APPLICANT: Fernandes, Elma  
APPLICANT: Vernet, Corine  
APPLICANT: Yang, Meijia  
APPLICANT: Boldog, Ferenc  
APPLICANT: Herrmann, John  
TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby  
FILE REFERENCE: 15966-554 Cura-54 CON-S8  
CURRENT APPLICATION NUMBER: US/10/002, 304  
PRIOR FILING DATE: 2001-11-02  
PRIOR APPLICATION NUMBER: 09/604, 286  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 60/140, 584  
PRIOR FILING DATE: 1999-06-23  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 2155  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (166)..(1935)  
US-10-002-304-11

Query Match  
Best Local Similarity 82.2%; Score 14.8; DB 9; Length 2155;  
Best Local Similarity 88.9%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18  
|||||  
DB 1298 TGGCGCCGATGCCAGTTG 1281

RESULT 10  
US-10-003-152-11/c  
Sequence 11, Application US/10003152  
Patent No. US20020151494A1  
GENERAL INFORMATION:  
APPLICANT: Shimkets, Richard  
APPLICANT: Fernandes, Elma  
APPLICANT: Vernet, Corine  
APPLICANT: Yang, Meijia  
APPLICANT: Boldog, Ferenc  
APPLICANT: Herrmann, John  
TITLE OF INVENTION: No. US20020151494A1 Amino Acid Sequences for Human Semaphorin  
FILE REFERENCE: 15966-554 Cura-54 CON-S12  
CURRENT APPLICATION NUMBER: US/10/003, 152  
PRIOR FILING DATE: 2001-11-02  
PRIOR APPLICATION NUMBER: 09/604, 286  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 60/140, 584  
PRIOR FILING DATE: 1999-06-23  
NUMBER OF SEQ ID NOS: 49

```

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2155
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(1935)
US-10-003-152-11

Query Match      82.2%; Score 14.8; DB 12; Length 2155;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGCGCGCTTGCCCAATTG 18
        ||||||| ||||| |||
Db      1298 TGGCGCGCATGCCAGTTG 1281

RESULT 11
US-10-002-050-21/c
; Sequence 21, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corline
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20030032095A1 Nucleic Acid Sequences Encoding Human Semaph
; FILE REFERENCE: 15966-554 Cura-54 CON-S14
; CURRENT APPLICATION NUMBER: US/10/002,050
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 21
; LENGTH: 2156
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(2037)
; NAME/KEY: variation
; LOCATION: (1)..(2156)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-050-21

Query Match      82.2%; Score 14.8; DB 9; Length 2156;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGCGCGCTTGCCCAATTG 18
        ||||||| ||||| |||
Db      1298 TGGCGCGCATGCCAGTTG 1281

RESULT 12
US-10-002-304-21/c
; Sequence 21, Application US/10002304
; Publication No. US20030036185A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corline
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
```

```

; FILE REFERENCE: 15966-554 Cura-54 CON-S8
; CURRENT APPLICATION NUMBER: US/10/002,304
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 21
; LENGTH: 2156
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(2037)
; NAME/KEY: variation
; LOCATION: (1)..(2156)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-21

Query Match      82.2%; Score 14.8; DB 12; Length 2156;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGCGCGCTTGCCCAATTG 18
        ||||||| ||||| |||
Db      1298 TGGCGCGCATGCCAGTTG 1281

RESULT 13
US-10-003-152-21/c
; Sequence 21, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corline
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1 Amino Acid Sequences for Human Semaphorin
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 21
; LENGTH: 2156
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(2037)
; NAME/KEY: variation
; LOCATION: (1)..(2156)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-21

Query Match      82.2%; Score 14.8; DB 12; Length 2156;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGCGCGCTTGCCCAATTG 18
        ||||||| ||||| |||
Db      1298 TGGCGCGCATGCCAGTTG 1281

RESULT 14
```



US-10-002-050-13/c  
; Sequence 13, Application US/10002050  
; Publication No. US20030032095A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Semaphorin  
; FILE REFERENCE: 15966-554 Cura-54 CON-S14  
; CURRENT APPLICATION NUMBER: US/10/002,050  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 13  
; LENGTH: 2284  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (166)..(1953)  
; NAME/KEY: variation  
; LOCATION: (1)..(2284)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-050-13

Query Match 82.2%; Score 14.8; DB 9; Length 2284;  
Best Local Similarity 88.9%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18  
||||||| ||||| |||  
DB 1298 TGGCGCCGATGCCAGTTG 1281

RESULT 15  
US-10-002-304-13/c  
; Sequence 13, Application US/10002304  
; Publication No. US20030036185A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby  
; FILE REFERENCE: 15966-554 Cura-54 CON-S8  
; CURRENT APPLICATION NUMBER: US/10/002,304  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 13  
; LENGTH: 2284  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (166)..(1953)  
; NAME/KEY: variation  
; LOCATION: (1)..(2284)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-304-13

Query Match 82.2%; Score 14.8; DB 9; Length 2284;  
Best Local Similarity 88.9%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18  
||||||| ||||| |||  
DB 1298 TGGCGCCGATGCCAGTTG 1281

Search completed: June 21, 2003, 00:34:38  
Job time : 8.65094 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 19:31:16 ; Search time 2520.75 Seconds

(without alignments)  
6927.165 Million cell updates/sec

Title: US-09-965-553-11

Perfect score: 600

Sequence: 1 ttggagcgtgcgcataata.....tccatagccacaagatctca 600

Scoring table: IDENTITY\_NUC

Gapped 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_cm:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_on:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rtd:\*

36: em\_htg\_mam:\*

37: em\_htg\_vrt:\*

38: em\_sy:\*

39: em\_htgo\_hum:\*

40: em\_htgo\_mus:\*

41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	571.2	95.2	8573	8	AF186184
2	532	88.7	10128	8	AF186182
3	529.6	88.3	762	8	AF378068
4	526.4	87.7	762	8	AF378062
5	523.2	87.2	12886	8	AF186183
6	502.6	83.8	763	8	AF378063
7	430	71.7	762	8	AF378064
8	359	59.8	92281	8	AP004896
9	352.6	58.8	762	8	AF378072
10	351.6	58.6	762	8	AF378070
11	349.8	58.3	761	8	AF378073
12	348.8	58.1	761	8	AF378069
13	346.2	57.7	110729	2	AC123573
14	345.6	57.6	13637	8	AF186185
15	342.2	57.0	761	8	AF378071
16	340.4	56.7	762	8	AF378065
17	334	55.7	762	8	AF378067
18	332.6	55.4	761	8	AF378066
19	330.2	55.0	762	8	AF378049
20	327	54.5	762	8	AF378050
21	323.8	54.0	3147	8	AB007466
22	323.2	53.9	264461	2	AC130804
23	319.6	53.3	80930	8	AP004894
24	319	53.2	762	8	AF378051
25	319	53.2	2531	8	PSJ000639
26	315.8	52.6	12314	8	PSJ000640
27	313.6	52.3	760	8	AF378015
28	311.6	51.9	108720	8	AB046436
29	311.6	51.9	138940	8	AC063973
30	308.8	51.5	87814	8	AB046426
31	308.8	51.5	52616	2	AC090029
32	308.4	51.4	762	8	AF378057
33	308	51.3	60482	8	AC090030
34	307.8	51.3	18813	8	T9E19
35	307.8	51.3	35313	8	AB028613
36	307.8	51.3	38519	6	AX059548
37	307.8	51.3	48422	6	AX059509
38	307.8	51.3	94805	8	F2112
39	307.8	51.3	196766	8	ATCHR1Y16
40	307.2	51.2	85962	8	AB046431
41	307.2	51.2	95463	8	F1809
42	306.2	51.0	762	8	AF378079
43	306.2	51.0	108827	8	ATF7P3
44	306.2	51.0	108720	8	AB046436
45	306.2	51.0	138940	8	AC063973

# ALIGNMENTS

RESULT 1

AF186184

LOCUS AF186184

DEFINITION Glycine max retrovirus-like element Calypso3-1, partial sequence.

ACCESSION AF186184

VERSION AF186184.1 GI:6671124

KEYWORDS

SOURCE

ORGANISM

Glycine max.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

1 (bases 1 to 8573)

Wright,D.A. and Voytas,D.F.

Pred. No. is the number of results predicted by chance to have a

TITLE Calypso: A Heterogeneous Retrovirus-like Element Family from Glycine max

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 8573)  
AUTHORS Wright, D.A. and Voytas, D.F.  
TITLE Direct Submission  
JOURNAL Submitted (14-SEP-1999) Zoology and Genetics, Iowa State University, 2208 Molecular Biology Bldg., Ames, IA 50011, USA

FEATURES  
source  
1. 8573  
/organism="Glycine max"  
/strain="L85"  
/db\_xref="taxon:3847"  
repeat\_region  
1. 8452  
/rpl\_family="retrovirus-like element Calypso3-1"  
/rpl\_type="dispersed"  
1. 3781  
/gene="pol"  
/pseudo  
4461. 5722  
/gene="envelope-like"  
/pseudo  
7034. 7039  
misc\_feature  
/note="polyurine tract"  
7043. 8452  
LTR  
BASE COUNT 2584 a 1731 c 2009 g 2249 t  
ORIGIN

Query Match 95.28; Score 571.2; DB 8; Length 8573;  
Best Local Similarity 97.0%; Pred. No. 8.6e-172;  
Matches 582; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

1 TTGGAGGCTGGGCTCATATACCCCATCTGACAGCGCTGGTAAGCCGATACAGG 60  
1172 TTGGAGGCTGGGCTCATATACCCCATCTGACAGCGCTGGTAAGCCGATACAGG 1231  
61 GTTCCCAAGAAAGGTGAATGACAGTGTACAGATGAGAGAAATGACTTGATACCA 120  
1232 GTTCCCAAGAAAGGTGAATGACAGTGTACAAATGAGAGAAATGACTTGATACCA 1291  
121 CGAAGTGTCTACTGGTGGCAATGTATGACATATGCCAAGCTGAATGAAGCCACGG 180  
1292 CGAAGTGTCTACTGGTGGCAATGTATGACATATGCCAAGCTGAATGAAGCCACGG 1351  
181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGAT 240  
1352 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGAT 1411  
241 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
1412 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1471  
301 CAGAGAGACGCGCTTACATGCGCCCTTGGCGCTTGGCTTACAGAGAGATGCGATTC 360  
1472 CAGAGAGACGCGCTTACATGCGCCCTTGGCGCTTGGCTTACAGAGAGATGCGATTC 1531  
361 GGGTTATGATGACACGACCATTTTCAGAGTGCATGCTGCGCATTTTTCAGACATG 420  
1532 GGGTTATGATGACACGACCATTTTCAGAGTGCATGCTGCGCATTTTTCAGACATG 1591  
421 GTGGAGAAAAGCATGAGATATTTATGAGACATTTCTGGCTTTTGGACCCCATTTGAC 480  
1592 GTGGAGAAAAGCATGAGATATTTATGAGACATTTCTGGCTTTTGGACCCCATTTGAC 1651  
481 AGCTGTTGAGAACTAGAGAGGTACTTACAGAGTGCAGAGAGATGAGATGAGATG 540  
1652 AGCTGTTGAGAACTAGAGAGGTACTTACAGAGTGCAGAGAGATGAGATGAGATG 1711  
541 AATTGGAAAAGTGTCAATTTTCATGCTGAGAGGCGATAGTCTTAGGCGACAAAGATCTCA 600  
1712 AATTGGAAAAGTGTCAATTTTCATGCTGAGAGGCGATAGTCTTAGGCGACAAAGATCTCA 1771

RESULT 2  
AF186182 10128 bp DNA linear PLN 05-JAN-2000

LOCUS AF186182  
DEFINITION Glycine max retrovirus-like element Calypsol-1, partial sequence.  
ACCESSION AF186182  
VERSION AF186182.1 GI:6671122  
KEYWORDS  
SOURCE  
ORGANISM  
Glycine max.  
Glycine max.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 10128)  
AUTHORS Wright, D.A. and Voytas, D.F.  
TITLE Calypso: A Heterogeneous Retrovirus-like Element Family from Glycine max

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 10128)  
AUTHORS Wright, D.A. and Voytas, D.F.  
TITLE Direct Submission  
JOURNAL Submitted (14-SEP-1999) Zoology and Genetics, Iowa State University, 2208 Molecular Biology Bldg., Ames, IA 50011, USA

FEATURES  
source  
1. 10128  
/organism="Glycine max"  
/strain="L85"  
/db\_xref="taxon:3847"  
repeat\_region  
294. >10128  
/rpl\_family="retrovirus-like element Calypsol-1"  
/rpl\_type="dispersed"  
294. 1598  
1605. 1623  
1759. 7172  
primer\_bind  
1759. 7172  
/gene="pol"  
/pseudo  
8068. 8976  
/gene="envelope-like"  
/pseudo

BASE COUNT 3031 a 2135 c 2330 g 2632 t  
ORIGIN

Query Match 88.78; Score 532; DB 8; Length 10128;  
Best Local Similarity 94.0%; Pred. No. 3.3e-159;  
Matches 564; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

1 TTGGAGGCTGGGCTCATATACCCCATCTGACAGCGCTGGTAAGCCGATACAGG 60  
4564 TTGGAGGCTGGGCTCATATACCCCATCTGACAGCGCTGGTAAGCCGATACAGG 4623  
61 GTTCCCAAGAAAGGTGAATGACAGTGTACAGATGAGAGAAATGACTTGATACCA 120  
4624 GTTCCCAAGAAAGGTGAATGACAGTGTACAGATGAGAGAAATGACTTGATACCA 4683  
121 CGAAGTGTCTACTGGTGGCAATGTATGACATATGCCAAGCTGAATGAAGCCACGG 180  
4684 CGAAGTGTCTACTGGTGGCAATGTATGACATATGCCAAGCTGAATGAAGCCACGG 4743  
181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGAT 240  
4744 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGAT 4803  
241 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
4804 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4862  
301 CAGAGAGACGCGCTTACATGCGCCCTTGGCGCTTGGCTTACAGAGAGATGCGATTC 360  
4863 CAGAGAGACGCGCTTACATGCGCCCTTGGCGCTTGGCTTACAGAGAGATGCGATTC 4922  
361 GGGTTATGATGACACGACCATTTTCAGAGTGCATGCTGCGCATTTTTCAGACATG 420  
4923 GGGTTATGATGACACGACCATTTTCAGAGTGCATGCTGCGCATTTTTCAGACATG 4982

QY 421 GTGAGAAAGCATGAGGTATTTATGACGACACTTCGCTTTTGGACCCCATTTGAC 480  
 DB 4983 GTGGAGAGAGATGAGGTATTTATGACGACACTTCGCTTTTGGACCCCATTTGAC 5042  
 QY 481 AGCTTTTGGAGAACCTTAGAGAGGCTTCTTCAAGAGTGGCAAGACTAATTGGTACTG 540  
 DB 5043 AACTTTTGGAGAACCTTAGAGAGTGGCTTCTTCAAGAGTGGCAAGACTAATTGGTACTG 5102  
 QY 541 AATTGGGAAAGTGCATTTTCATGCTTGGAGAGGAGGATGCTCTAGGCCCAAGATCTCA 600  
 DB 5103 AATTGGGAAAGTGCATTTTCATGCTTGGAGAGGAGGATGCTCTAGGCCCAAGATCTCA 5162  
 RESULT 3  
 AF378068  
 LOCUS Glycine max isolate Soybean8-3 retrovirus element Calypso, partial  
 DEFINITION sequence.  
 ACCESSION AF378068  
 VERSION AF378068.1 GI:15724032  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Glycine max.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.  
 REFERENCE 1 (bases 1 to 762)  
 AUTHORS Wright,D.A. and Voytas,D.F.  
 TITLE Athila4 of Arabidopsis and Calypso of soybean define a lineage of  
 endogenous plant retroviruses  
 JOURNAL Genome Res. 12 (1), 122-131 (2002)  
 MEDLINE 21638318  
 PUBMED 11779837  
 REFERENCE 2 (bases 1 to 762)  
 AUTHORS Voytas,D.F. and Wright,D.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa  
 State University, 2208 Molecular Biology, Ames, IA 50011, USA  
 FEATURES  
 source  
 1..762  
 /organism="Glycine max"  
 /variety="Williams"  
 /isolate="Soybean8-3"  
 /db\_xref="taxon:3847"  
 /note="endogenous\_virus: Calypso"  
 <1..>762  
 /note="nonfunctional reverse transcriptase protein due to  
 mutation"  
 misc\_feature  
 BASE COUNT 217 a 145 c 200 g 200 t  
 ORIGIN  
 Query Match 88.3%; Score 529.6; DB 8; Length 762;  
 Best Local Similarity 92.7%; Pred. No. 1.6e-158;  
 Matches 556; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
 QY 1 TTGGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAGCCCATGACAGTG 60  
 DB 25 TTGGAGGCTGGGCTATATATACCTCATCTCTGACAGCGCTTGGTAGCCCATGACAGTG 84  
 QY 61 GTTCCAGAGAAAGTGAATGACAGTGTACAGATGAGAGAGAAATGACTGAAACCAACA 120  
 DB 85 GCTCCAGAGAAATGAGATGACAGTGTACAGATGAGAGAGAAATGACTGAAACCAACA 144  
 QY 121 CGAAGTGTACCTGTTGGCGAATGTGTATGCACTATCGAAGCTGAAATGAGCAACGCG 180  
 DB 145 CGAAGTGTACCTGTTGGCGAATGTGTATGCACTATCGAAGCTGAAATGAGCAACGCG 204  
 QY 181 AAGGACCAATTCCTTACCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 240  
 DB 205 AAGGACCAATTCCTTACCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 264

QY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
 DB 265 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 324  
 QY 301 CAGGAGAAAGCGGCTTTTACATGCCCCCTTTGGCGCTTTTGTCTTACAGAAAGATGCCATTG 360  
 DB 325 CAGGAGAAAGATGCGCTTTTACATGCCCCCTTTGGCGCTTTTGTCTTACAGAAAGATGCCATTG 384  
 QY 361 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
 DB 385 AGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 444  
 QY 421 GTGAGAAAGCATGAGGTATTTATGACGACACTTCGCTTTTGGACCCCATTTGAC 480  
 DB 445 GTGAGAGAGCATGAGGTATTTATGACGACACTTCGCTTTTGGACCCCATTTGAC 504  
 QY 481 AGCTTTTGGAGAACCTTAGAGAGGCTTCTTCAAGAGTGGCAAGACTAATTGGTACTG 540  
 DB 505 AGTTCCTTAAAGCACTTAAAGAGTGTACTACAGAGTGGCTTAAAGACTAATTGGTACTA 564  
 QY 541 AATTGGGAAAGTGCATTTTCATGCTTGGAGAGGAGGATGCTCTAGGCCCAAGATCTCA 600  
 DB 565 AATTGGGAAAGTGCATTTTCATGCTTGGAGAGGAGGATGCTCTAGGCCCAAGATCTCA 624  
 RESULT 4  
 AF378062  
 LOCUS Glycine max isolate Soybean8-2 retrovirus element Calypso, partial  
 DEFINITION sequence.  
 ACCESSION AF378062  
 VERSION AF378062.1 GI:15724025  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Glycine max.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.  
 REFERENCE 1 (bases 1 to 762)  
 AUTHORS Wright,D.A. and Voytas,D.F.  
 TITLE Athila4 of Arabidopsis and Calypso of soybean define a lineage of  
 endogenous plant retroviruses  
 JOURNAL Genome Res. 12 (1), 122-131 (2002)  
 MEDLINE 21638318  
 PUBMED 11779837  
 REFERENCE 2 (bases 1 to 762)  
 AUTHORS Voytas,D.F. and Wright,D.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa  
 State University, 2208 Molecular Biology, Ames, IA 50011, USA  
 FEATURES  
 source  
 1..762  
 /organism="Glycine max"  
 /variety="L85"  
 /isolate="Soybean8-2"  
 /db\_xref="taxon:3847"  
 /note="endogenous\_virus: Calypso"  
 <1..>762  
 /note="nonfunctional reverse transcriptase protein due to  
 mutation"  
 misc\_feature  
 BASE COUNT 213 a 145 c 203 g 201 t  
 ORIGIN  
 Query Match 87.7%; Score 526.4; DB 8; Length 762;  
 Best Local Similarity 92.3%; Pred. No. 1.7e-157;  
 Matches 554; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
 QY 1 TTGGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAGCCCATGACAGTG 60  
 DB 25 TTGGAGGCTGGGCTATATATACCTCATCTCTGACAGCGCTTGGTAGCCCATGACAGTG 84  
 QY 61 GTTCCAGAGAAAGTGAATGACAGTGTACAGATGAGAGAGAAATGACTGAAACCAACA 120

gene	7909 . 9171	/gene="envelope-like"
misc_feature	11223 . 11228	/pseudo
LTR	11232 . 12569	/note="polypurine tract"
BASE COUNT	3882 a 2456 c 2767 g .3781 t	
ORIGIN		
Query Match	87.2%; Score 523.2; DB 8; Length 12886;	
Best Local Similarity	93.2%; Ped. No.2.3e-156;	
Matches 559; Conservative	0; Mismatches 38; Indels 3; Gaps 1	
OY	1 TTGAGGGCTGGGCTCATATATACCCCATTCCTGACAGCCGCTGGGCTAAAGCCCACTACAGGTG 60	
DB	4293 TTGGAGGGCTGGGCTCATATATACCCCTTCCTTAACAGCTGCTGGGTAAAGCCCACTACAGGTG 4355	
OY	61 GTTCCCAAGAAAGGTGGAATGACAGCTGTGACAGATGAGAGAAATGATCTGATTACCAACA 120	
DB	4353 GTTCCCAAGAAAGGTGGAATGACAGCTGTGACAGAAATGAGAAATGATCTGATTACCAAGA 4411	
OY	121 CGAATGTCACCTGCTGTGGCGAATGTGTATGACTATGCCAAGCTAATGAAAGCCACAGG 180	
DB	4413 CGAATATACACTGCTGTGGCGAATGTGTATCAACTATGCCAAGCTAATGAAAGCCACAGA 4472	
OY	181 AAGACACATTTCCCTTACCTTTACATGCGCTTTCATGATCAGATGCTGAGAGACTTGACAGGCA 240	
DB	4473 AAGACACATTTCCCTTACCTTTACATGCGCTTTCATGATCAGATGCTGAGAGACTTGAGGCA 4533	
OY	241 TACTACTTTCTTGGATGGATACCTCGGGATATCAACCAAGTCGCGGTAGACCCCAAGAT 300	
DB	4533 TACTACTTTCTTGGATGGATACCTCGGGATATTAATCAAGATCGCGGTAGACCCCAAGAT 4592	
OY	301 CAGGAGAAAGAGGCGCTTTACATGCGCTTTGGCGCTTTCGTTACAGAAAGATGGCATTC 360	
DB	4593 CAAGAGAAAGGCGGCTTTACATGCGCTTTGGCGCTTTCGTTATAGAAAGATGGCATTC 4655	
OY	361 GGGTTATGTATGCACACGACATTTAGAGGTGATGCTGCGCATTTTTCACAGATG 420	
DB	4653 GGGTTATGTATGCACACGACATTTAGAGGTGATGCTGCGCATTTTTCACAGATG 4712	
OY	421 GTGGAGAAACATCGAGGTATTATATGACGACTTCGCGGTTTTGGAGCCCTCATTTGAC 480	
DB	4713 GTTGAAGAAACATCGAGGTATTATATGACGACTTCGCGGTTTTGGAGCCCTCATTTAAC 4773	
OY	481 AGCTGTTTGAGGAACCTTAGAGAGGCTACTCAGAGGTGCGGAAGACTTAACCTGTACTG 540	
DB	4773 A---GTTGAGGAACCTTAGAGGTACTTATAGAGTTGATGATAGAGACTTAACCTGTACTG 4829	
OY	541 AATTGGGAAAAGCTCATTTTATGTGTTGAGAGGGCATATGTCCTTAGGCAACAAAGTCTCA 600	
DB	4830 AACTGGGAGAAGTGTCACTTATGCTTCMAAGAGGGCATGCTTAGGCAACAAAGTCTCA 4889	
RESULT 6		
AF378063	763 bp	DNA linear PLN 11-JAN-2002
LOCUS	AF378063	
DEFINITION	Glycine max isolate Soybean2 retrovirus element Calypso, partial sequence.	
ACCESSION	AF378063	
VERSION	AF378063.1	
KEYWORDS	GI:15724026	
SOURCE		
ORGANISM	Glycine max.	
	Glycine max.	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.	
REFERENCE	1 (bases 1 to 763)	
AUTHORS	Wright,D.A. and Voytas,D.F.	
TITLE	Attitud of Arabidopsis and Calypso define a lineage of endogenous plant retroviruses	
JOURNAL	Genome Res. 12 (1), 122-131 (2002)	

MEDLINE	216398318
PUBMED	11779837
REFERENCE	2 (bases 1 to 763)
AUTHORS	Voytas,D.F. and Wright,D.A.
TITLE	Direct Submission
JOURNAL	Submitted (08-MAY-2001) Department of zoology and Genetics, Iowa State University, 2208 Molecular Biology, Ames, IA 50011, USA
FEATURES	
source	Location/Qualifiers 1..763 /organism="Glycine max" /variety="L85" /isolate="Soybean2" /db_xref="taxon:3847" /note="endogenous_virus: Calypso"
misc-feature	<1..>763 /note="nonfunctional reverse transcriptase protein due to mutation"
BASE COUNT	213 a      143 c      203 g      204 t
ORIGIN	
Query Match	83.8%; Score 502.6; DB 8; Length 763;
Best Local Similarity	90.8%; Pred.No. 7.5e-150;
Matches 546; Conservative	0; Mismatches 54; Indels 1; Gaps 1
OY	1 TTGGAGGTGGGGCTCATATTACCCATCTCTGACAGCGTTGGGTAGGCCACTACAGTG 60
Db	 25 TTGGAGGTGGGGCTCATATTACCCATCTCTAATGACCACTTGGTAGGCCACTACAGTG 84
OY	61 GTTCCCAGAAGAAGTGAATGACAGTGGTAGAGATGAGAGGAATGACTTGATTACCACA 120
Db	85 GTTCCCAGAAGAAGTGGAAATACACTAGTACGAATGGAAGAATGACCTTGATTACCACA 144
OY	121 CGAACGTGCTAGTGGTGGCGAATGTGATGCATATCGCAAGCTGAATGAAGCCACACGG 180
Db	145 CGAACTGTCATAGCTGGCGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 204
OY	181 AAGAGCACCTTCCCCCTTACCTTTACATGATGATGATGATGATGATGATGATGATGATGAT 240
Db	205 AAGAGCACCTTCCCCCTTACCTTTACATGATGATGATGATGATGATGATGATGATGATGAT 264
OY	241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db	265 TATTATGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324
OY	301 CAGAGGAAGACGCGCCTTTACATGCC -CTTTGGCGCTTTGCTTACAGAGAGTGCATT 359
Db	325 CAAAGAGAAGACGACCTTCACATGCCCTTTTGGCGCTTTGCTTACAGAGAGTGCATT 384
OY	360 CGGGTTATGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 419
Db	385 CGGGTTATGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444
OY	420 GGTTGAGAAAAGCATCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 479
Db	445 GGTTGAGAAAAGCATCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 504
OY	480 CAGCTGTTTGAAGAACCTAGAGAGGTAATTCAAGAGTGCAGAGAGTGAATCTTGTTACT 539
Db	505 CAGTTGTTTGAAGAACCTAGAGAGTGAATCTTAGAGTGCGCTAGAGAGTGAATCTTAG 564
OY	540 GAATTGGGAAAAGTGTCTATTCTATGATGATGATGATGATGATGATGATGATGATGATGAT 599
Db	565 GAATGTTGGGAAAAGTGTCTATTCTATGATGATGATGATGATGATGATGATGATGATGATGAT 624
OY	600 A 600
Db	625 A 625
RESULT 7	
LOCUS	AF378064                762 bp        DNA                linear        PLN 11-JAN-2002
DEFINITION	Glycine max isolate Soybean9-2 retrovirus element Calypso reverse

Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Medline	PubMed	Reference	Authors	Title	Journal	Features	Source
transcriptase gene, partial cds.																
AF378064	1	GI:15724027														
glycine max.																
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoidae; Phaseoleae; Glycine.																
1 (bases 1 to 762)																
Wright, D.A. and Voytas, D.F.																
Attila4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses																
Genome Res. 12 (1), 122-131 (2002)																
21638318																
1179837																
2 (bases 1 to 762)																
Voytas, D.F. and Wright, D.A.																
Direct Submission																
Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa State University, 2208 Molecular Biology, Ames, IA 50011, USA																
Location/Qualifiers																
1..762																
/organism="Glycine max"																
/variety="L85"																
/isolate="Soybean9-2"																
/db_xref="taxon:3847"																
/note="endogenous_virus: Calypso"																
<1..>762																
/product="reverse transcriptase"																
<1..>762																
/codon_start=1																
/product="reverse transcriptase"																
/protein_id="AI06417.1"																
/db_xref="GI:15724028"																
/translation="VREVKLLEVLGIIDPIDSANVSPVLVSKRGMTVIREKND LIPRVTSMKLCIDYRKLEAKRKHFPPEFDMOERLGAHYCFDAPGYNOI LVPRDDEKMAFCPEGFVARRIRPFICNAATPFOMCMALFADIVLEKIEVMD SVFEPSESCIKLEWLFORCEVNTLNEKHPFVWRBESIVLGHKISTGIEVDGK IDVIERKPEPSNVKGIKIFSLGQARFRRFKDPTKV"																
BASE COUNT	228 a	150 c	181 g	203 t												
ORIGIN																
Query Match	71.7%	Score 430;	DB 8;	Length 762;												
Best Local Similarity	82.4%	Pred. No. 1.5e-126;														
Matches	493;	Conservative	0;	Mismatches	105;	Indels	0;	Gaps	0;							
2	TGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGGTAGACCCAGTACAGGTG	61														
26	TAGAGTTGGGCTTATCTACCCATCTCCAGACAGCGCTTGGTAAAGCCAGTCTGGTGG	85														

```

Db      386 GGTGTGCAATGACCACTACCATTCCAAATGTCATGTGGCCATTTTTCAGATATAG 445
OY      422 TGGAGAAAGCATGAGATATTTATGAGACCTTCGGTTTGTGACCCCATTTGACA 481
Db      446 TGGAGAAAGCATGAGATATTTATGAGACCTTCGGTTTGTGACCCCATTTGACA 505
OY      482 GCTGTTTGGAGAACCTAGAGAGGTACTTTCAGAGGTGGAGAGACTAATGTTACTGA 541
Db      506 GTTGTGGAAGACTGTGATGATGATACAAAGATGGTGGAGAAACAATAGTACTA 565
OY      542 ATTGGAGAAAGTGTATTCATGTTTCAGAGAGGATAGTCTAGGCCCAAGATGTC 599
Db      566 ATTGGAGAAAGTGTATTCATGTTTCAGAGAGGATAGTCTAGGCCCAATTAATTC 623

RESULT 8
AP004896 92281 bp DNA linear PLN 19-JUL-2002
LOCUS     Lotus japonicus genomic DNA, chromosome 5, clone:LjT16107, TM0040,
DEFINITION complete sequence.
ACCESSION AP004896
VERSION   AP004896.1 GI:21907912
KEYWORDS  HG.
SOURCE    Lotus japonicus DNA, clone_LjT library clone:LjT16107.
ORGANISM  Lotus japonicus
REFERENCE 1 Kameko,T., Nakamura,Y., Asamizu,E., Kato,T., Sato,S. and Tabata,S.
AUTHORS   Structural Analysis of a Lotus japonicus Genome. I. Sequence
TITLE      Features and Mapping of Sixty-six TAC clones which cover the 6.7 Mb
JOURNAL    Regions of the Genome
REFERENCE 2 (bases 1 to 92281)
AUTHORS   Nakamura,Y.
TITLE      Direct Submission
JOURNAL    Submitted (26-MAR-2002) Yasukazu Nakamura, Kazusa DNA Research
           Institute, Department of Plant Gene Research, 1532-3, Yana,
           Kisarazu, Chiba 292-0812, Japan (E-mail: yn@kazusa.or.jp,
           URL: http://www.kazusa.or.jp, Tel:81-438-52-3935,
           Fax:81-438-52-3934)
FEATURES   Location/Qualifiers
            source          1..92281
                        /organism="Lotus japonicus"
                        /db_xref="taxon:34305"
                        /chromosome="5"
                        /clone="LjT16107"
                        /clone_lib="LjT library"
                        /note="TAC clone:TM0040"
BASE COUNT 29650 a 16291 c 17416 g 28924 t
ORIGIN
Query Match 59.8%; Score 359; DB 8; Length 92281;
Best Local Similarity 75.0%; Pred. No. 1.3e-103;
Matches 445; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

OY      2 TGGAGGTGGGCTCATATACCCATCTGACAGCGCTGGGTAGGCCAGTACAGTGG 61
Db      78712 TAGAGCAGGTATGATTTATCCATTTCTGATGATGTTCTGGGTGATCCAGTGG 78771
OY      62 TTCCCAAGAAAGTGAATGACAGTGTACGAGATGAGAGAAATGACTTACCAAC 121
Db      78772 TTCCAAAGAAAGAGAGCATGACTGTGATTAATGACAGAAATGAGTGATTCGACAA 78831
OY      122 GAAGTGTCACTGTTGGCGAATGTGTATGACTATGCAAGTGAATGAAGCCACACGA 181
Db      78832 GAACAGTGTCTAGTTGGAGAAATGTGTATGATTAAGAGAACTGAAACAGCCACACGAA 78891
OY      182 AGGACATTTCCCTTACCTTTCATGATGATGCTGAGAGAGCTGAGAGGCGAGCAT 241

```

```

Db      78892 AGGACATTTCCCTTACCATTCATGATCAATGATGTTGGAGAGATTTACTGGGAGCCCT 78951
OY      242 ACTACTGTTCTTGGATGATATCTGGGATACAAACAGATCGCGGTAGACCACAGATC 301
Db      78952 TCTATTCCTTCTTGATGATATCTGGGCTACAAATTAATGATCAACCCAGAAAGCC 79011
OY      302 AGGAGAAAGCGCCCTTTACATGCCCCCTTGGCGTCTTGTTCACAGAAAGATGCCATTG 361
Db      79012 AGGAGAAAGCGCTTTCACATGCTCCCTTGGTGTTCATAGGAAATGCGCATTTGG 79071
OY      362 GGTTATGTAATGACAGCCCATTTTCAGAGGTGATGCGCCATTTTTCAGACATGG 421
Db      79072 GGTGTATATGCTCCAGCAACCTTTCAGATGATCTTCGATATTTTCAGATTTGG 79131
OY      422 TGGAGAAAGCATGAGATATTTATGAGAGCTTCTCGTTTTCGACCCCATTTGACA 481
Db      79132 TGGAGAGATCATGAGATTTGATGATGATTTCTAGTATTTGGTCTCTTGTGAT 79191
OY      482 GCTGTTTGGAGAACCTAGAGAGGTACTTTCAGAGTGGGAGAGACTAATGTTGATC 541
Db      79192 TATGTTGCAAAACCTTGACACCGGTGGAAGCGGTGTGGAACCAATCTGCTTCTTA 79251
OY      542 ATTGGAGAAAGTGTATTCATGTTTCGAGAGGCGATAGTCTAGGCCCAAGATCTCA 600
Db      79252 ATTGGAGAAAGTGTATTCATGTTTCGAGAGGCGATAGTCTAGGCCCAATCTCA 79310

RESULT 9
AF378072 762 bp DNA linear PLN 11-JAN-2002
LOCUS     Glycine max isolate soybean5-1 retrovirus element Calypso, partial
DEFINITION sequence.
ACCESSION AF378072
VERSION   AF378072.1 GI:15724037
KEYWORDS  Glycine max.
SOURCE    Glycine max.
ORGANISM  Glycine max.
REFERENCE 1 Wright,D.A. and Voytas,D.F.
AUTHORS   Athila4 of Arabidopsis and Calypso of soybean define a lineage of
TITLE      endogenous plant retroviruses
JOURNAL    Genome Res. 12 (1), 122-131 (2002)
MEDLINE   21638318
PUBMED    11779837
REFERENCE 2 (bases 1 to 762)
AUTHORS   Voytas,D.F. and Wright,D.A.
TITLE      Direct Submission
JOURNAL    Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa
           State University, 2208 Molecular Biology, Ames, IA 50011, USA
FEATURES   Location/Qualifiers
            source          1..762
                        /organism="Glycine max"
                        /variety="Hark"
                        /isolate="Soybeans-1"
                        /db_xref="taxon:3847"
                        /note="endogenous_virus: Calypso"
misc_feature <1..>762
            /note="nonfunctional reverse transcriptase protein due to
            mutation"
BASE COUNT 224 a 132 c 186 g 220 t
ORIGIN
Query Match 58.8%; Score 352.6; DB 8; Length 762;
Best Local Similarity 74.3%; Pred. No. 1e-101;
Matches 445; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

OY      2 TGGAGGTGGGCTCATATACCCATCTGACAGCGCTGGGTAGGCCAGTACAGTGG 61
Db      26 TGGAGCAGGCTTATTTATTCATTTGCGATAGTGGGTTAGCCCTGTGACAGTTG 85

```



Qy	62	TTCCCAAGAAAGTGGAAATGACAGTGGTAAAGAGATGAGAGAAATGACATGCTTATGCCAACAC	121		
Db	86	TTCCCAAGAAAGTGGAAATGACAGTGGTAAAGAGATGAGAGAAATGACATGCTTATGCCAAC	145		
Qy	122	GAACTGTACAGTGGTGGCGAATGTGTATGCAGTATGCCAGTATGCCAATGAAATGAAACACACGA	181		
Db	146	GGACTATACACCGGGTGGAGATGTGTATGCTATGCAGTATGCCAATGAAATGAAATGAAACAC	205		
Qy	182	AGAGCAATTCCTCCCTTACCTTTTCATGATCAGATGCTGGAGAGACCTTGCAGAGCAGGCAT	241		
Db	206	AGGACCAATTCACCTCCCTTTCATGATCAGTATGCCAATGAAATGCTGAAAGACCTTGCAGAG	265		
Qy	242	ACTACCTGTTTCCTTGATGATGATATCAGTATGCCAGTATGCCAATGAAATGCTGAAAGAC	301		
Db	266	ATTATGTTGTTTTCCTTGATGATGATATCAGTATGCCAGTATGCCAATGAAATGCTGAAAG	325		
Qy	302	AGAGAAAGACGCCCTTATACATGCCCTTTGGCGCTCTTGTCTTACAGAAAGATGCCATTCG	361		
Db	326	AAGAGAGACATGCTTACACCTTATCCCTTTGGTATTCCTATTCGCTATTCGCGGCGAGCCCT	385		
Qy	362	GCTATGTATATGACACACGACCATTTTCAGAGTGTGATGCTGCGCATTTTTCACAGCATGG	421		
Db	386	GTTTGTCCAAAGCCCGCCAGCTACATTTTCAGAGTGTATGATATACATATTTTCTCATATAGG	445		
Qy	422	TGAGAAAGACATGAGAGTATTTATTTGAGCAGCTTCGCTTGGTTTGGACCCATTCATTCACA	481		
Db	446	TGGAATAATGATTTGAAAGTTTTCATGAGCAGATGTCTCATTTTGGGCGCATCTTTTGAAG	505		
Qy	482	GCTGTTTGGAGAACCTAGAGAGGGTACTTCAGAGGTGCGAAGAGACTAATCTGTTACTGA	541		
Db	506	GTGCTTATATCAAAATCTTGAAGAAGTATTTAAAGACGTAAGAGAGTCCAAATGATCTCA	565		
Qy	542	ATTGGGAAAGTGCATTTTCATGCTTGGTGGAGAGGCGATATGCTTGGCCACAAAGATCTCA	600		
Db	566	ATTGGGAAAGTGCATTTTCATGCTTGGTGGAGAGGCGATATGCTTGGCCATTAATTTTCA	624		
RESULT 10	AF378070	762 bp	DNA	linear	PLN 11-JAN-2002
LOCUS	AF378070				
DEFINITION	Glycine max isolate Soybean3 retrovirus element Calypso reverse transcriptase gene, partial cds.				
ACCESSION	AF378070				
VERSION	AF378070.1	GI:15724034			
KEYWORDS					
SOURCE	Glycine max.				
ORGANISM	Glycine max.				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.				
AUTHORS	1 (bases 1 to 762)				
TITLE	Wright, D.A. and Voytas, D.F.				
JOURNAL	Athlia4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses				
MEDLINE	Genome Res. 12 (1), 122-131 (2002)				
PUBMED	21638318				
REFERENCE	11779837				
AUTHORS	2 (bases 1 to 762)				
TITLE	Voytas, D.F. and Wright, D.A.				
JOURNAL	Direct Submission				
FEATURES	Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa State University, 2208 Molecular Biology, Ames, IA 50011, USA				
SOURCE	Location/Qualifiers				
	1..762				
	/organism="Glycine max"				
	/variety="Williams"				
	/isolate="Soybean3"				
	/db_xref="taxon:3847"				
	/note="endogenous_virus: Calypso"				
	<1..>762				
	/product="reverse transcriptase"				
mRNA					

CDS  
 <1..762  
 /codon\_start=1  
 /product="reverse transcriptase"  
 /protein\_id="AAB06418.1"  
 /db\_xref="GI:15724035"  
 /translation="VRKEVEKLEBAGLITPISDSSWVSPOVYVRKKGATVVKNDRE  
 LIPRRVTRMNCIDYRKLNEATRKRDHYPIPEFMOYMKRLAROSFYEFELGYSYNOI  
 AVPDQDEKTAETCFSEVFARMRPEFICNASTFFQCCMAIFPDMEKCLEYMDPEF  
 SFEGASRGNCLANEKYLORCEKSNLILNMEKHFMOVEGIVLGHKISKRIEIVYKREK  
 LDVIDRKPPPNVNVGHISFLGHGFFYRFRIFKPKV"  
 BASE COUNT 239 a 125 c 168 g 230 t  
 ORIGIN  
 Query Match 58.6%; Score 351.6; DB 8; Length 762;  
 Best Local Similarity 74.2%; Pred. No. 2.2e-101;  
 Matches 444; Conservative 0; Mismatches 154; Indels 0; Gaps 0;  
 Oy 2 TGGAGCGCTGGGCTCATATACCCCATCTCTGACAGCGCTTGGGTAAACCCAGTACAGGTGG 61  
 Db 26 TAGAGGACGAGGCGCTCATATACCCCAATTCAGATACCTCTCGGTGTAAGTCGAGTGG 85  
 Oy 62 TTCCCAAGAAAGGTGGATGACAGTGGATGACAGATGAGAGGAATGACTTGTATACCAAC 121  
 Db 86 TTCCCAAAAAAGAGGAGATGACAGTGGTAAAAAATGATAGAAATAGAGTATTTCTTACAA 145  
 Oy 122 GAAGTGTACTGTGTGGCGAATGTGTATCGACTATCCGACGTAATGAAGCAACACGCA 181  
 Db 146 GAAGAGTCACAGATGAGATGTGTATTTGATTATAGAGAACCTCAATGAGCCACAAGAA 205  
 Oy 182 AGGACCAATTCGCCCTTACCTTCAATGGATACAGATCGAGAGAGACTGGACAGGAGCAT 241  
 Db 206 AAGACCATATCCCACTTCCTTCACAGATCAAAATGTATAGAGACTTGCAAGCAAGATCT 265  
 Oy 242 ACTACTGTTCTTGGATGGATATCTCGGGATACAAACAGATGCGGTAGACCCAGAGATC 301  
 Db 266 TCTACCGTTCTTGGACGAGATATCATAGGTTACAAATCAGATGTGCAATGGATCTCGAGATC 325  
 Oy 302 AGGAGAAAGCGCCCTTATCATGCCCCCTTTGGCGTCTTGGCTTACAGAAAGATGGCATTCG 361  
 Db 326 AAGAAAAAAGCTTTTATCATGCTCTTCAATGTGTTTGGTTATGCGCCGACCGCGTTCG 385  
 Oy 362 GGTATGTAAATGACACACAGCCACATTTTCAGAGGTGCATCTGCCCATTTTTCACAGATGG 421  
 Db 386 GTTATGTAAATGCCCTCATCTACTTTTTCAGAGATGATATGAGCAATTTTGTATGACATGG 445  
 Oy 422 TGGAGAAAAGCATGAGGTATTTATGACGCACTTCGCGTTTGGACCCCATTTTGACA 481  
 Db 446 TAGAGAAATGTATGAACTTTTATGATGATGATTTTCTCTTCTTGGTGCACTTTTGGAA 505  
 Oy 482 GCTGTTTAGAGACCTTAGAGAGGGTACTTACAGAGTGGAGAGAGACTACTTGGATAGCA 541  
 Db 506 ATTGCTTAGCAAAATTTAGAGAAAGTGTTCACAACTGTGAAAAATCTAATTTGGTGGCTTA 565  
 Oy 542 ATTGGGAAAAGTGTCAATTTTCATGGTTGCGAGAGGGCATATGCTTAGAGCCACAAGATCTC 599  
 Db 566 ACTGGGAAAATGTCACTTTATGTGATACAAAGAGTATGTGCTTAGAGACACAAAATCTC 623  
 RESULT 11  
 LOCUS AF378073 761 bp DNA linear PLN 11-JAN-2002  
 DEFINITION Glycine max isolate Soybean5 retrovirus element Calypso, partial  
 sequence.  
 ACCESSION AF378073  
 VERSION AF378073.1 GI:15724038  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Glycine max.  
 Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.  
 REFERENCE  
 1 (bases 1 to 761)

AUTHORS Wright, D.A. and Voytas, D.F.  
 TITLE Athlia4 of Arabidopsis and Calypso of soybean define a lineage of  
 endogenous plant retroviruses  
 JOURNAL Genome Res. 12 (1), 122-131 (2002)  
 MEDLINE 21638318  
 PUBMED 11779837  
 REFERENCE 2 (bases 1 to 761)  
 AUTHORS Voytas, D.F. and Wright, D.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa  
 State University, 2208 Molecular Biology, Ames, IA 50011, USA

FEATURES  
 source  
 1..761  
 /organism="Glycine max"  
 /variety="Hark"  
 /isolate="Soybeans"  
 /db\_xref="taxon:3847"  
 /note="endogenous\_virus: Calypso"  
 misc\_feature  
 <1..>761  
 /note="nonfunctional reverse transcriptase protein due to  
 mutation"

BASE COUNT 217 a 130 c 185 g 229 t  
 ORIGIN

Query Match 58.3%; Score 349.8; DB 8; Length 761;  
 Best Local Similarity 75.2%; Pred. No. 8.3e-101;  
 Matches 449; Conservative 0; Mismatches 147; Indels 1; Gaps 1;

OY 2 TGGAGGCGGGCTCATATACCCCATCTGACAGCGCTGGGTAGCCGACATACAGTGG 61  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db TGGAGGAGGCGGCTCATCTTCTTATCTCAGATAGTG-TTGGGTGAGTCAGTGCATGG 84  
 OY 62 TTCCCAAGAAAGGTGATGATGACAGTGTACGAGATGAGAGAACTTGATACCAAC 121  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db TTCCCAAGAAAGGTGAGAACTGTGTGAGAAATAGAAATAGCCATCTTCAACC 144  
 OY 85 TTCCCAAGAAAGGTGAGAACTGTGTGAGAAATAGAAATAGCCATCTTCAACC 144  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 122 GAAGTGTACAGTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGATG 181  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db GAAGTGTACAGTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGATG 204  
 OY 145 GAAGTGTACAGTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGATG 204  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 182 AGGACATTTCCCTTACCTTACATGATGATGATGATGATGATGATGATGATGATGATG 241  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db AGGATCACTTCCCTTACCTTACATGATGATGATGATGATGATGATGATGATGATGATG 264  
 OY 205 AGGATCACTTCCCTTACCTTACATGATGATGATGATGATGATGATGATGATGATGATG 264  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 242 ACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db TCTATTATTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 324  
 OY 302 AGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 361  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db AAGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 384  
 OY 325 AAGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 384  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 362 GGTATATTAATGACAGCCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATG 421  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db AACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 444  
 OY 422 TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 481  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 504  
 OY 445 TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 504  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 482 GCTGTTTGGAGAACTGAGAGGCTTACATGATGATGATGATGATGATGATGATGATGATG 541  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db GCTGTTTGGAGAACTGAGAGGCTTACATGATGATGATGATGATGATGATGATGATGATG 564  
 OY 542 ATTGGAGAAAGTGTATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 598  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db ATTGGAGAAAGTGTATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 621  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12  
 AF378069  
 LOCUS AF378069 761 bp DNA linear PLN 11-JAN-2002  
 DEFINITION Glycine max isolate Soybean2 retrovirus element Calypso, partial

sequence.  
 AF378069  
 VERSION AF378069.1 GI:15724033  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Glycine max.  
 Glycine max.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.

REFERENCE 1 (bases 1 to 761)  
 AUTHORS Wright, D.A. and Voytas, D.F.  
 TITLE Athlia4 of Arabidopsis and Calypso of soybean define a lineage of  
 endogenous plant retroviruses  
 JOURNAL Genome Res. 12 (1), 122-131 (2002)  
 MEDLINE 21638318  
 PUBMED 11779837  
 REFERENCE 2 (bases 1 to 761)  
 AUTHORS Voytas, D.F. and Wright, D.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa  
 State University, 2208 Molecular Biology, Ames, IA 50011, USA

FEATURES  
 source  
 1..761  
 /organism="Glycine max"  
 /variety="Williams"  
 /isolate="Soybean2"  
 /db\_xref="taxon:3847"  
 /note="endogenous\_virus: Calypso"  
 misc\_feature  
 <1..>761  
 /note="nonfunctional reverse transcriptase protein due to  
 mutation"

BASE COUNT 223 a 131 c 187 g 220 t  
 ORIGIN

Query Match 58.1%; Score 348.8; DB 8; Length 761;  
 Best Local Similarity 74.3%; Pred. No. 1.7e-100;  
 Matches 440; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

OY 2 TGGAGGCGGGCTCATATACCCCATCTGACAGCGCTGGGTAGCCGACATACAGTGG 61  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db TGGAGGAGGCGGCTCATTTTATTCATTTCCATTTGCGATGTGATGATGATGATGATGATG 85  
 OY 62 TTCCCAAGAAAGGTGATGATGACAGTGTACGAGATGAGAGAACTTGATACCAAC 121  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db TTCCCAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 145  
 OY 86 TTCCCAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 145  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 122 GAAGTGTACAGTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGATGATG 181  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db GAAGTGTACAGTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGATGATG 205  
 OY 146 GAAGTGTACAGTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGATGATG 205  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 182 AGGACATTTCCCTTACCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 241  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db AGGATCACTTCCCTTACCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 265  
 OY 206 AGGATCACTTCCCTTACCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 265  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 242 ACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db ATTATTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 325  
 OY 266 ATTATTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 325  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 302 AGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 361  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db AAGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 385  
 OY 326 AAGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 385  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 362 GGTATATTAATGACAGCCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATG 421  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db GGTATATTAATGACAGCCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATG 445  
 OY 386 GTTTGTCATTTGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 445  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 505  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 422 TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 481  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 505  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 OY 482 GCTGTTTGGAGAACTGAGAGGCTTACATGATGATGATGATGATGATGATGATGATGATG 541  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||



TITLE Calypso: A Heterogeneous Retrovirus-like Element Family from Glycine max

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 13637)

AUTHORS Wright, D.A. and Voytas, D.F.

TITLE Direct Submission

JOURNAL Submitted (14-SEP-1999) Zoology and Genetics, Iowa State University, 2208 Molecular Biology Bldg., Ames, IA 50011, USA

FEATURES

SOURCE 1. 13637

Location/Qualifiers

organism="Glycine max"

strain="L85"

db\_xref="taxon:3847"

repeat\_region 3335..12472

3335..12472

/rpl\_family="retrovirus-like element Calypso-1"

gene 3335..7628

/gene="pol"

gene 8205..9350

/pseudo

gene 8205..9350

/gene="envelope-like"

/pseudo

misc\_feature 11100..11106

/note="Polyurine tract"

LTR 11109..12472

BASE COUNT 4227 a 2455 c 2912 g 4043 t

ORIGIN

Query Match 57.6%; Score 345.6; DB 8; Length 13637;

Best Local Similarity 73.5%; Pred. No. 2.3e-99;

Matches 441; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

1 TTGGAGGCTGGGCTCATATACCCCATCTCTGACAGCGCTTGAGTAAAGCCAGTACAGTG 60

4979 TAGAAGAGAGGCTTATTTACCCCTCCGATGATGATGATGATGATGATGATGATGATG 5038

61 GTTCCCAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120

5039 GTCCCAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5098

121 CGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180

5099 AGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5158

181 AAGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 240

5159 AAGGACATTTACCACTTCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 5218

241 TACTACTGTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

5219 TATTATGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5278

301 CAGGAGAAAGGCTTACATGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 360

5279 CAGGAGAAAGGCTTACATGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 5338

361 GGGTATATGATGACACGACATTTGAGAGTGCATGCTGCGCATTTTTCAGACATG 420

5339 GGGTATGATGACACGACATTTGAGAGTGCATGCTGCGCATTTTTCAGACATG 5398

421 GTGAGAAAGGCTGAGAGTATTTATGACGACCTTCTCGCTTTTTCGACCTTTCATTTGAC 480

5399 GTGAGAAAGGCTGAGAGTATTTATGACGACCTTCTCGCTTTTTCGACCTTTCATTTAAG 5458

481 AGCTGTTGAGAGCTGAGAGTATTTATGAGAGTGCATGCTGCGCATTTTTCAGACATG 540

5459 GGGTATATTAATCTTGAAGAGTATTTACAGAGTGCATGCTGCGCATTTTTCAGACATG 5518

541 AATTGAGAAAGGCTGAGAGTATTTATGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 600

5519 AATTGAGAAAGGCTGAGAGTATTTATGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 5578

RESULT 15

AF378071

LOCUS 761 bp DNA linear PLN 11-JAN-2002

DEFINITION Glycine max isolate Soybean2 retrovirus element Calypso, partial sequence.

ACCESSION AF378071

VERSION AF378071.1

KEYWORDS GI:15724036

SOURCE

ORGANISM

Glycine max.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae;

REFERENCE 1 (bases 1 to 761)

AUTHORS Wright, D.A. and Voytas, D.F.

TITLE Athila4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses

JOURNAL Genome Res. 12 (1), 122-131 (2002)

MEDLINE 21638318

PUBMED 11798337

REFERENCE 2 (bases 1 to 761)

AUTHORS Voytas, D.F. and Wright, D.A.

TITLE Direct Submission

JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa State University, 2208 Molecular Biology, Ames, IA 50011, USA

FEATURES

SOURCE 1. 761

Location/Qualifiers

organism="Glycine max"

variety="Hark"

isolate="Soybean2"

db\_xref="taxon:3847"

/note="endogenous\_virus: Calypso"

/note="nonfunctional reverse transcriptase protein due to mutation"

BASE COUNT 224 a 134 c 180 g 223 t

ORIGIN

Query Match 57.0%; Score 342.2; DB 8; Length 761;

Best Local Similarity 74.3%; Pred. No. 2.3e-98;

Matches 445; Conservative 0; Mismatches 153; Indels 1; Gaps 1;

2 TGAAGCTGGGCTCATATACCCCATCTCTGACAGCGCTTGAGTAAAGCCAGTACAGTG 61

26 TAGAAGAGAGCTTATTTATCCATTTGCGATGATGATGATGATGATGATGATGATGATGATG 85

62 TTCCCAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121

86 TCCCGAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 145

122 GAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181

146 GAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 204

182 AGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 241

205 AGGACATTTATTCACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 264

242 ACTACTGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301

265 ATTATGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324

302 AGGAGAAAGGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 361

325 AGGAGAAAGGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 384

362 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 421

385 GCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 444

422 TGAAGAAAGGCTGAGAGTATTTATGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 481

Sat Jun 21 17:38:36 2003

us-09-965-553-11.rge

Page 11

Db 445 TGGAAAAATGCATTGACATTTTTCATGGACGATTTTCATTTTTTGGGCACATCTTTGGAG 504

Oy 482 GCTGTTTGAGGAACTAGAGGGGTACTTCACAGGGCGCAAGACTACTGTGCTCTGA 541

Db 505 GCTACTTTCMAAAGCTTGAAAGAGTATTCACAGAGATGTGAGAGGTATTAATCTGTTCTCA 564

Oy 542 AFTGGGAAAGATGTATTTTCATGTGTTCCGAGAGGGCATGCTCTAGGCCACAGATCTCA 600

Db 565 ATTGGGACAAATAGCCATTTCATGTGTTCAAGAGGATATGTGCGGGCATTTAAATTTTCA 623

Search completed: June 20, 2003, 23:24:00  
Job time : 2522.09 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:46:16 ; Search time 343.396 Seconds  
(without alignments)  
3934.810 Million cell updates/sec

Title: US-09-965-553-11

Perfect score: 600

Sequence: 1 ttggagcgtgcgcataata.....tccatggccacaagatctca 600

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq\_101002:\*  
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	600	100.0	600	21	AAZ35258
2	600	100.0	12286	21	AAZ35261
3	571.2	95.2	4609	21	AAZ35273
4	568.2	94.7	597	21	AAZ35280
5	532	88.7	9829	21	AAZ35271
6	523.2	87.2	12571	21	AAZ35272
7	345.6	57.6	9139	21	AAZ35274
8	313.8	52.3	597	21	AAZ35266
9	307.8	51.3	94895	21	AAZ352302

10	307.8	51.3	1082138	21	AAF22305	Arabidopsis thaliana
11	304.6	50.8	600	21	AAZ35264	Arabidopsis retroe
12	304.6	50.8	59590	21	AAF22281	BAC containing rep
13	303	50.5	83590	21	AAF22283	BAC containing rep
14	303	50.5	90336	21	AAF22289	BAC containing rep
15	303	50.5	94895	21	AAF22302	BAC containing rep
16	303	50.5	103929	21	AAF22287	BAC containing rep
17	303	50.5	134499	21	AAF22286	BAC containing rep
18	303	50.5	1082138	21	AAF22305	Arabidopsis thaliana
19	301.4	50.2	79122	21	AAF22294	BAC containing rep
20	300	50.0	92584	21	AAF22288	BAC containing rep
21	295.8	49.3	129021	21	AAF22296	BAC containing rep
22	295	49.2	163319	21	AAF22306	Arabidopsis thaliana
23	291.4	48.6	64415	21	AAF22279	BAC containing rep
24	287.8	48.0	64415	21	AAF22279	BAC containing rep
25	287	47.8	109973	21	AAF22298	BAC containing rep
26	274.2	45.7	611590	21	AAF22203	Arabidopsis thaliana
27	273.4	45.6	40349	21	AAF22278	BAC containing rep
28	272	45.3	96988	21	AAF22290	BAC containing rep
29	251.6	41.9	80450	21	AAF22295	BAC containing rep
30	172.4	28.7	10482	21	AAZ35275	Soybean retroelme
31	159.2	26.5	109973	21	AAF22298	BAC containing rep
32	102.8	17.1	91552	24	AAZ38803	BAC clone K6P36 fr
33	97.2	16.2	94618	21	AAF22285	BAC containing rep
34	97	16.2	8435	20	AAZ3445	Genomic sequence o
35	91.4	15.2	82588	21	AAF22301	BAC containing rep
36	89.4	14.9	90336	21	AAF22289	BAC containing rep
37	89.4	14.9	121001	21	AAF22284	BAC containing rep
38	86	14.3	101786	21	AAF22293	BAC containing rep
39	85.6	14.3	59590	21	AAF22281	BAC containing rep
40	85.6	14.3	151826	21	AAF22291	BAC containing rep
41	84.2	14.0	4271	23	ABL04659	Drosophila melanog
42	84.2	14.0	6535	23	ABL04658	Drosophila melanog
43	82.2	13.7	163319	23	AAF22306	Arabidopsis thaliana
44	81	13.5	20213	23	ABL21212	Drosophila melanog
45	74.8	12.5	96583	21	AAF22297	BAC containing rep

#### ALIGNMENTS

RESULT 1	
AAZ35258	
ID	AAZ35258 standard; DNA; 600 BP.
AC	AAZ35258;
XX	
DT	27-MAR-2000 (first entry)
XX	
DE	Plant retroelment generic reverse transcriptase gene.
XX	
KW	Retroelment; retrovirus; transgenic plant; gene transfer;
KW	soybean; pea; reverse transcriptase; Calypso; athalia; cyclops; ss.
XX	
OS	Glycine max.
OS	Arabidopsis thaliana.
OS	Pisum sativum.
XX	
PN	WO960842-A2.
XX	
PD	02-DEC-1999.
XX	
PF	28-MAY-1999; 99WO-US11858.
XX	
PR	29-MAY-1998; 98US-0087125.
PR	28-MAY-1999; 99US-0087125.
PA	(WRIG/) WRIGHT D A.
PA	(VOYT/) VOYTAS D F.
XX	
PI	Wright DA, Voytas DF;
XX	
DR	WPI; 2000-105586/09.

DR P-PSDB; AAV32431.

XX New nucleic acid molecules for imparting agronomically significant  
PT characters to plants, especially soybean

PS Clatm 1(e); Page 80; 118pp; English.

CC The present sequence comprises a consensus of reverse transcriptase  
CC gene sequences identified in retrovirus-like element (retroelement)  
CC calypso of soybean (see AA235280), cyclops of pea (see AA235266) and  
CC athlia of Arabidopsis thaliana (see AA235264). The invention  
CC provides molecular tools in the form of retroelements and  
CC retroelement-containing vectors, cells and plants. Methods are  
CC provided for introducing the retroelements into cells, especially  
CC when the retroelement carries at least 1 agronomically-significant  
CC characteristic (ACS). In a preferred method, a helper cell line  
CC which expresses gag, pol and env sequences is used to enable  
CC retroelement sequences that allow for replication and integration.  
CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
CC sequence selected from a retroelement primer binding site,  
CC envelope, gag, integrase, reverse transcriptase, protease or  
CC RNase-H sequence (see AA235254-61). Also provided are plant  
CC retroviral particles that are used to transfer the nucleic acids  
CC into plant cells.

CC Sequence 600 BP; 157 A; 128 C; 166 G; 149 T; 0 other;

Query Match 100.0%; Score 600; DB 21; Length 600;

Best Local Similarity 100.0%; Pred. No. 6.5e-193;  
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAGCCAGTACAGTG 60  
DB 1 TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAGCCAGTACAGTG 60  
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGTAGAGATGAGAGAAATGACTTGTATCCACAA 120  
DB 61 GTTCCCAAGAAAGGTGGAATGACAGTGTAGAGATGAGAGAAATGACTTGTATCCACAA 120  
QY 121 CGAAGTGTACTGTGTGGCGAATGTATCGACTTCGCAAGCTGAATGAAACCCACAGG 180  
DB 121 CGAAGTGTACTGTGTGGCGAATGTATCGACTTCGCAAGCTGAATGAAACCCACAGG 180  
QY 181 AAGGACCATTTCCCTTACCTTTATGATAGATGCTGAGAGAGACTTGCAGGAGGCA 240  
DB 181 AAGGACCATTTCCCTTACCTTTATGATAGATGCTGAGAGAGACTTGCAGGAGGCA 240  
QY 241 TACTACTGTTTCTTGATGATGATCTCGGATACCAACGATCGCGTAGACCCCGAGAGAT 300  
DB 241 TACTACTGTTTCTTGATGATGATCTCGGATACCAACGATCGCGTAGACCCCGAGAGAT 300  
QY 301 CAGGAGAAAGAGCGGCTTTACATGCCCCCTTGGCGTCTTGGTTACAGAAAGATGCAATTC 360  
DB 301 CAGGAGAAAGAGCGGCTTTACATGCCCCCTTGGCGTCTTGGTTACAGAAAGATGCAATTC 360  
QY 361 GGGTTATGTATGACCAACCAATTCAGAGGTGATGCTGGCCATTTTTCAGACATG 420  
DB 361 GGGTTATGTATGACCAACCAATTCAGAGGTGATGCTGGCCATTTTTCAGACATG 420  
QY 421 GTTGAGAGAAAGCATCGAGTATTATGAGACATCTTCGGTTTGGACCCCTCATTTTAC 480  
DB 421 GTTGAGAGAAAGCATCGAGTATTATGAGACATCTTCGGTTTGGACCCCTCATTTTAC 480  
QY 481 AGCTGTTTGGAGAACTAGAGAGGCTACTTCAGAGGTGCGAAGAGACTAATCTGTACTG 540  
DB 481 AGCTGTTTGGAGAACTAGAGAGGCTACTTCAGAGGTGCGAAGAGACTAATCTGTACTG 540  
QY 541 AATTGGGAAAAGTGTATTTCAATGTTGAGAGGGCATATGCTTAGGCCACAGAATCTCA 600  
DB 541 AATTGGGAAAAGTGTATTTCAATGTTGAGAGGGCATATGCTTAGGCCACAGAATCTCA 600

RESULT 2

AA235261  
ID AA235261 standard; DNA; 12286 BP.

XX AA235261;

XX 27-MAR-2000 (first entry)

XX Plant generic retroelement.

XX Retroelement; retrovirus; transgenic plant; gene transfer;

XX soybean; pea; Calypso; athlia; cyclops; ss.

XX Glycine max.

XX Arabidopsis thaliana.

XX Pisum sativum.

XX Key

XX CDS

XX W09960842-A2.

XX 28-MAY-1999; 99W0-US11858.

XX 29-MAY-1998; 98US-0087125.

XX 28-MAY-1999; 99US-0087125.

XX (WRIG/) WRIGHT D A.

XX (VOYT/) VOYTAS D F.

XX Wright DA, Voytas DF;

XX WPI: 2000-105586/09.

XX P-PSDB; AAV32434.

XX New nucleic acid molecules for imparting agronomically significant

XX characters to plants, especially soybean

XX Claim 1(h); Page 84-88; 118pp; English.

XX The present sequence comprises a generic plant retroelement

XX obtained from retrovirus-like elements (retroelements) calypso of

XX soybean, cyclops of pea and athlia of Arabidopsis thaliana. The

XX invention provides molecular tools in the form of retroelements and

XX retroelement-containing vectors, cells and plants. Methods are

XX provided for introducing the retroelements into cells, especially

XX when the retroelement carries at least 1 agronomically-significant

XX characteristic (ACS). In a preferred method, a helper cell line

XX which expresses gag, pol and env sequences is used to enable

XX retroelement sequences that allow for replication and integration.

XX Claimed isolated nucleic acid molecules comprise a nucleic acid

XX sequence selected from a retroelement primer binding site,

XX envelope, gag, integrase, reverse transcriptase, protease or

XX RNase-H sequence (see AA235254-61). Also provided are plant

XX retroviral particles that are used to transfer the nucleic acids

XX into plant cells.

SQ Sequence 12286 BP; 3748 A; 2540 C; 2767 G; 3331 T; 0 other;

Query Match 100.0%; Score 600; DB 21; Length 12286;

Best Local Similarity 100.0%; Pred. No. 3.5e-192;  
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAGCCAGTACAGTG 60  
DB 4278 TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAGCCAGTACAGTG 4337  
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGTAGAGATGAGAGAAATGACTTGTATCCACAA 120  
DB 61 GTTCCCAAGAAAGGTGGAATGACAGTGTAGAGATGAGAGAAATGACTTGTATCCACAA 120



Db	4338	GTTCGCCAAGAAAGGTGGAAATGACAGTGGTACAGATGTACAGATGTAGAGGAAATGACTTGATATCCACAA	4337
OY	121	CGAATCTCACTCGTGTGGCGAATGTGTATCATCTATTCGACATTCGCAAGCTGAATGAAGCCACACGG	180
Db	4398	CGAACTGTCACTCGTGTGGCGAATGTGTATCATCTATTCGACATTCGCAAGCTGAATGAAGCCACACGG	4457
OY	181	AAGACCATTTCCCTTACCTTTCATGTATCAGATTCAGATTCGTGGAGAGACTTTGCAAGGGCAGGCA	240
Db	4458	AAGACCATTTCCCTTACCTTTCATGTATCAGATTCAGATTCGTGGAGAGACTTTGCAAGGGCAGGCA	4517
OY	241	TACTACTGTTTCTTGGATGATACCTGGGATACACACAGATTCGGGTAGACCCACAGAT	300
Db	4518	TACTACTGTTTCTTGGATGATACCTGGGATACACACAGATTCGGGTAGACCCACAGAT	4577
OY	301	CAGAGAAAGCGGCTTATCATGCCCCCTTGGGGCTCTTGCCTTACAAAGATGCCATTC	360
Db	4578	CAGAGAAAGCGGCTTATCATGCCCCCTTGGGGCTCTTGCCTTACAAAGATGCCATTC	4637
OY	361	GGGTTATGTATGCACACGACCATTTTCAGAGGTGATGCTGGCCATTTTTTCAGACATG	420
Db	4638	GGGTTATGTATGCACACGACCATTTTCAGAGGTGATGCTGGCCATTTTTTCAGACATG	4697
OY	421	GTGGAGAAAGCAATCGAGGTATTATGTAGAGACTTCGCTTTTGGACCTCATTTGAC	480
Db	4698	GTGGAGAAAGCAATCGAGGTATTATGTAGAGACTTCGCTTTTGGACCTCATTTGAC	4757
OY	481	AGCTGTTTGAAGCACTTAGAGAGGTAATTCAGAGGTGCGCAAGAGACTTAATCTTGTAAGT	540
Db	4758	AGCTGTTTGAAGCACTTAGAGAGGTAATTCAGAGGTGCGCAAGAGACTTAATCTTGTAAGT	4817
OY	541	AATTGGGAAAAGTGATTTTCATAGTGTTCGAGAGGGCATAGTCCTAGAGCCACAAATCTCA	600
Db	4818	AATTGGGAAAAGTGATTTTCATAGTGTTCGAGAGGGCATAGTCCTAGAGCCACAAATCTCA	4877

RESULT 3	
ID	AAZ35273 standard; DNA; 4609 BP.
XX	AAZ35273;
AC	
DT	27-MAR-2000 (first entry)
XX	
DE	Soybean retroelement Calypso 1-3.
XX	
KW	Retroelement; retrovirus; transgenic plant; gene transfer;
KM	Calypso 1-3; soybean; ss.
XX	
OS	Glycine max.
XX	
PN	W09960842-A2.
XX	
PD	02-DEC-1999.
XX	
PF	28-MAY-1999; 99MO-US11858.
XX	
PR	29-MAY-1998; 98US-0087125.
XX	
PR	28-MAY-1999; 99US-0087125.
XX	
PA	(WRIG/) WRIGHT D A.
PA	(VOYT/) VOYTAS D F.
PI	Wright DA, Voytas DF;
XX	
DR	WPI; 2000-105586/09.
XX	
PT	New nucleic acid molecules for imparting agronomically significant
XX	characters to plants, especially soybean
XX	
PS	Example 3; Page 102-104; 110pp; English.

This is the nucleotide sequence of the Calypso 1-3 retroelement of soybean. It was identified by screening of a soybean lambda ccosybean.

CC library using a reverse transcriptase probe. 2 Groups of soybean  
CC retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3  
CC (see AMZ35271-73) and Calypso 2-1 and 2-2 (see AMZ35274-75). The  
CC retroelements include gag, pol, env and primer binding site  
CC sequences that can be used in constructs of the invention. The  
CC invention provides molecular tools in the form of retroelements and  
CC retroelement-containing vectors, cells and plants. Methods are  
CC provided for introducing the retroelements into cells, especially  
CC when the retroelement carries at least 1 agronomically significant  
CC characteristic (ACS). In a preferred method, a helper cell line  
CC which expresses gag, pol and env sequences is used to enable  
CC transfer of a secondary construct which carries an ACS and has  
CC retroelement sequences that allow for replication and integration.  
CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
CC sequence selected from a retroelement primer binding site, envelope,  
CC gag, integrase, reverse transcriptase, protease or RNase-H sequence  
CC (see AMZ3554-61). Also provided are plant retroviral particles that  
CC particles are used to transfer the nucleic acids into plant cells.  
XX  
SQ Sequence 4609 BP; 1420 A; 949 C; 1146 G; 1094 T; 0 other;

Query Match	95.2%	Score 571.2	DB 21	Length 4609
Best Local Similarity	97.0%	Pred. No. 1.2e-182		
Matches 582	Conservative 0	Mismatches 18	Indels 0	Gaps 0
Qy	1	TTGAGAGCGGCGCTATATATACCCATCTCGATGACAGCGCTGGGTGAAGCCAGTACAGTGTG	60	
Db	1174	TTGGAGGTTGGGCTCATATACCCATCTCTGACACAGCTTGGGTGAAGCCAGTACAGTGTG	1233	
Qy	61	GTTCCCAAGAAAGGTGGAATGACAGTGTGATACGAGATGAGAGGAATGACTTGATACCAACA	120	
Db	1234	GTTCCCAAGAAAGGTGGAATGACAGTGTGATACCAAAATGAGAGGAATGACTTGATACCAACA	1293	
Qy	121	CGAAGTGTACACTGTTGTGGCGAATGTATGACATATCGCAAGCTGAATGAAGCCACACGG	180	
Db	1294	CGAAGTGTACACTGTTGTGGCGAATGTATGACATATCACAAAGCTGAATGAAGCTACACGG	1353	
Qy	181	AAGGACACATTTCCCTTACCTTTCATGATGACATGCTGGAGAGACTTGGAGGGCAGGCA	240	
Db	1354	AAGGACACATTTCCCTTACCTTTCATGATGACATGCTGGAGAGACTTGGAGGGCAGGCA	1413	
Qy	241	TACTACTGTTCTTGGATGGATCTCGGGATACAAACAGATCGCGGTAGACCCAGAGAT	300	
Db	1414	TACTACTGTTCTTGGATGGATCTCGGGATACAAACAGATCGCGGTAGACCCAGATGAT	1473	
Qy	301	CAGGAGAAAGACGGCCTTTACATGCCCCCTTGGCGCTCTTCTTACAGAAAGATGCCATTC	360	
Db	1474	CAGGAGAAAGACGGCTTTACATGCCCCCTTGGCGCTCTTCTTACAGAAAGATGTCATTC	1533	
Qy	361	GGGTATGTATGATGACACACCCACATTTCAAGAGTGCATGCTGGCCATTTTTTCAGACATG	420	
Db	1534	GGGTATGTATGATGACACCCACATTTCAAGAGTGCATGCTGGCCATTTTTTCAGACATG	1593	
Qy	421	GTTGGAGAAAGATCGAGTATTTATGAGACATTTCTCGGTTTTTTGGACCCCTCATTTTGAC	480	
Db	1594	GTTGGAGAAAGATCGAGTATTTATGAGACATTTCTCGGTTTTTTGGACCCCTCATTTTGAC	1653	
Qy	481	AGCTGTTTGAGAACCTAGAGAGGTACTTCACAGAGTGCACAGAGAGATTAAGTGTGACTG	540	
Db	1654	AGCTGTTTGAGAACCTAGAGAGGTACTTCACAGAGTGCCTTAGAGACTTAAGTGTGACTG	1713	
Qy	541	AATTGGGAAAAGTGTCAATTTCAATGTTTCGAGAGGGCATTAAGTCCATAGGCCCAAGATCTCA	600	
Db	1714	AATTGGGAAAAGTGTCAATTTCAATGTTTCGAGAGGGCATTAAGTCCATAGGCCCAAGATCTCA	1773	
RESULT 4				
ID	AAZ35280			
XX	AAZ35280	standard; DNA; 597 BP.		
AC	AAZ35280;			
DT	27-MAR-2000	(first entry)		

```

XX  Soybean retroelement calypso reverse transcriptase gene.
DE  Retroelement; retrovirus; transgenic plant; gene transfer:
KM  reverse transcriptase gene; calypso; soybean; ss.
XX  Glycine max.
OS  WO9960842-A2.
PN  02-DEC-1999.
XX  28-MAY-1999; 99MO-US11858.
XX  29-MAY-1998; 98US-0087125.
PR  28-MAY-1999; 99US-0087125.
XX  (WRIGHT) WRIGHT D A.
PA  (VOYT/) VOYTAS D F.
XX  Wright DA, Voytas DF;
PI  WPI; 2000-105586/09.
XX  New nucleic acid molecules for imparting agronomically significant
PT  characters to plants, especially soybean
XX  Disclosure; Page 115; 118pp; English.
PS  The present sequence represents the reverse transcriptase gene of
XX  retrovirus-like element (retroelement) calypso of soybean. The
CC  invention provides molecular tools in the form of retroelements and
CC  retroelement-containing vectors, cells and plants. Methods are
CC  provided for introducing the retroelements into cells, especially
CC  when the retroelement carries at least 1 agronomically-significant
CC  characteristic (ACS). In a preferred method, a helper cell line
CC  which expresses gag, pol and env sequences is used to enable
CC  transfer of a secondary construct which carries an ASC and has
CC  retroelement sequences that allow for replication and integration.
CC  claimed isolated nucleic acid molecules comprise a nucleic acid
CC  sequence selected from a retroelement primer binding site, envelope,
CC  gag, integrase, reverse transcriptase, protease or RNase-H sequence
CC  (see AA23524-61). Also provided are plant retroviral particles that
CC  comprise a retrovirus protein encoded by a nucleic acid sequence
CC  encoding a plant retroviral envelope protein and a nucleic acid
CC  including e.g. the present sequence. These plant retroviral
CC  particles are used to transfer the nucleic acids into plant cells.
XX  Sequence 597 BP; 162 A; 121 C; 158 G; 156 T; 0 other:
SQ
Query Match 94.7%; Score 568.2; DB 21; Length 597;
Best Local Similarity 97.0%; Pred. No. 4e-182;
Matches 579; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1 TTGGAGGCTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGGTAAAGCCAGTACAGTGG 60
DB 1 TTGGAGGCTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGGTAAAGCCAGTACAGTGG 60
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGTAGAGATGAGAGAAATGACATGTTATACCAACA 120
DB 61 GTTCCCAAGAAAGGTGGAATGACAGTGTAGAGATGAGAGAAATGACATGTTATACCAACA 120
QY 121 CGAAGTGTCACTGTGGCGCAATGTGTATCGACTATCGCAAGCTGAATGACCCACAGCG 180
DB 121 CGAAGTGTCACTGTGGCGCAATGTGTATCGACTATCGCAAGCTGAATGACCCACAGCG 180
QY 181 AAGGACCATTTCCCTTACCTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 181 AAGGACCATTTCCCTTACCTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

```

```

QY 301 CAGAGAGACGCGCTTTATCATGCCCCCTTGGGCTCTTGTCTTACAGAGATGCCATTC 360
DB 301 CAGAGAGACGCGCTTTATCATGCCCCCTTGGGCTCTTGTCTTACAGAGATGCCATTC 360
QY 361 GGGTATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 361 GGGTATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
QY 421 GTGAGAGAAAGCAATCGAGGTATTTATGAGAGACTTCTCGGTTTTTGGACCCCTATTGAC 480
DB 421 GTGAGAGAAAGCAATCGAGGTATTTATGAGAGACTTCTCGGTTTTTGGACCCCTATTGAC 480
QY 481 AGCTGTTTGAAGAACCTAGAGAGGGTACTTACAGAGGTGCGAAGAGACTAACTTGGTACTG 540
DB 481 AGCTGTTTGAAGAACCTAGAGAGGGTACTTACAGAGGTGCGAAGAGACTAACTTGGTACTG 540
QY 541 AATTGGGAAAAAGTGTCAATTTTCATGTTTCGAGAGGGCATATGCTTACAGCCACAAATC 597
DB 541 AATTGGGAAAAAGTGTCAATTTTCATGTTTCGAGAGGGCATATGCTTACAGCCACAAATC 597

RESULT 5
AA235271
ID AA235271 standard; DNA; 9829 BP.
XX
AC AA235271;
XX
DT 27-MAR-2000 (first entry)
XX
DE Soybean retroelement Calypso 1-1.
XX
KM Retroelement; retrovirus; transgenic plant; gene transfer;
XX  Calypso 1-1; soybean; ss.
XX
OS Glycine max.
XX
PN WO9960842-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99MO-US11858.
XX
PR 29-MAY-1998; 98US-0087125.
XX  28-MAY-1999; 99US-0087125.
XX
PA (WRIGHT) WRIGHT D A.
XX  (VOYT/) VOYTAS D F.
XX
PI Wright DA, Voytas DF;
XX
DR WPI; 2000-105586/09.
XX
PT New nucleic acid molecules for imparting agronomically significant
XX  characters to plants, especially soybean
XX
PS Example 3; Page 95-98; 118pp; English.
XX
This is the nucleotide sequence of the Calypso 1-1 retroelement of
CC  soybean. It was identified by screening of a soybean lambda
CC  library using a reverse transcriptase probe. 2 Groups of soybean
CC  retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3
CC  (see AA235271-73) and Calypso 2-1 and 2-2 (see AA235274-75). The
CC  retroelements include gag, pol, env and primer binding site
CC  sequences that can be used in constructs of the invention. The
CC  invention provides molecular tools in the form of retroelements and
CC  retroelement-containing vectors, cells and plants. Methods are
CC  provided for introducing the retroelements into cells, especially
CC  when the retroelement carries at least 1 agronomically-significant
CC  characteristic (ACS). In a preferred method, a helper cell line
CC  which expresses gag, pol and env sequences is used to enable
CC  transfer of a secondary construct which carries an ASC and has
CC  retroelement sequences that allow for replication and integration.

```

CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
 CC sequence selected from a retroelement primer binding site, envelope,  
 CC gag, integrase, reverse transcriptase, protease or RNase-H sequence,  
 CC (see AA235274-61). Also provided are plant retroviral particles that  
 CC particles are used to transfer the nucleic acids into plant cells.

Sequence 9829 BP; 2948 A; 2067 C; 2260 G; 2554 T; 0 other;

Query Match 88.7%; Score 532; DB 21; Length 9829;  
 Best Local Similarity 94.0%; Pred. No. 3.7e-169;  
 Matches 564; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

QY 1 TTGGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTGGGTAAGCCAGTACAGTG 60  
 DB 4271 TTGGAGGCTGGGCTCATATACCCATCTCTGATAGCGCTGGGTAAGTCCAGTACAGTG 4330  
 QY 61 GTTCCCAAGAAAGGTGATGACAGTGTGATGACAGATGAGAGATGATGATACCAACA 120  
 DB 4331 GTTCCCAAGAAAGGTGATGACAGTGTGATGACAGATGAGAGATGATGATACCAACA 4390  
 QY 121 CGAAGTGTCTGTTGGGATGTGTATGACTATCGCAAGCTGATGAAGCCACACG 180  
 DB 4391 CGAAGTGTCTGTTGGGATGTGTATGACTATCGCAAGTGAATGAAGCCACACG 4450  
 QY 181 AAGGACATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 DB 4451 AAGGACATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 4510  
 QY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 DB 4511 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4569  
 QY 301 CAGGGAAGACGGCTTACATGCCCCCTTGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 360  
 DB 4570 CAGGGAAGACGGCTTACATGCCCCCTTGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4629  
 QY 361 GGGTATGTAATGACACACCATTTTCAAGAGTGTGATGATGATGATGATGATGATGATGATGAT 420  
 DB 4630 GGGTATGTAATGACACACCATTTTCAAGAGTGTGATGATGATGATGATGATGATGATGATGAT 4669  
 QY 421 GTGGAAGAAAGCATGAGTATTTATGACGACTTCTCGGTTTGGACCCCTCAATTTGAC 480  
 DB 4690 GTGGAAGAAAGCATGAGTATTTATGACGACTTCTCGGTTTGGACCCCTCAATTTGAC 4749  
 QY 481 AGCTGTTTGAAGACCTTAAGAGAGTACTTCAAGAGTGTGATGATGATGATGATGATGATGATGAT 540  
 DB 4750 AGCTGTTTGAAGACCTTAAGAGAGTACTTCAAGAGTGTGATGATGATGATGATGATGATGATGAT 4809  
 QY 541 AATTGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 DB 4810 AATTGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4869

RESULT 6  
 AA235272  
 ID AA235272 standard; DNA; 12571 BP.

AC AA235272;  
 DT 27-MAR-2000 (first entry)  
 DE Soybean retroelement Calypso 1-2.

KM Retroelement; retrovirus; transgenic plant; gene transfer;  
 KM Calypso 1-2; soybean; ss.

OS Glycine max.

PN WO960842-A2.

PD 02-DEC-1999.

PF 28-MAY-1999; 99WO-US11858.

XX 29-MAY-1998; 98US-0087125.  
 PR 28-MAY-1999; 99US-0087125.  
 XX  
 PA (WRIG/) WRIGHT D A.  
 PA (VOYT/) VOYTAS D F.  
 PI Wright DA, Voytas DF.  
 DR WPI: 2000-105586/09.  
 XX  
 PT New nucleic acid molecules for imparting agronomically significant  
 PS characters to plants, especially soybean

Example 3; Page 98-102; 118pp; English.

CC This is the nucleotide sequence of the Calypso 1-2 retroelement of  
 CC soybean. It was identified by screening of a soybean lambda  
 CC library using a reverse transcriptase probe. 2 groups of soybean  
 CC retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3  
 CC (see AA235271-73) and Calypso 2-1 and 2-2 (see AA235274-75). The  
 CC retroelements include gag, pol, env and primer binding site  
 CC sequences that can be used in constructs of the invention. The  
 CC invention provides molecular tools in the form of retroelements and  
 CC retroelement-containing vectors, cells and plants. Methods are  
 CC provided for introducing the retroelements into cells, especially  
 CC when the retroelement carries at least 1 agronomically-significant  
 CC characteristic (ACS). In a preferred method, a helper cell line  
 CC which expresses gag, pol and env sequences is used to enable  
 CC transfer of a secondary construct which carries an ACS and has  
 CC retroelement sequences that allow for replication and integration.  
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
 CC sequence selected from a retroelement primer binding site, envelope,  
 CC gag, integrase, reverse transcriptase, protease or RNase-H sequence  
 CC (see AA235274-61). Also provided are plant retroviral particles that  
 CC particles are used to transfer the nucleic acids into plant cells.

Sequence 12571 BP; 3801 A; 2382 C; 2707 G; 3661 T; 0 other;

Query Match 87.2%; Score 523.2; DB 21; Length 12571;  
 Best Local Similarity 93.2%; Pred. No. 4.1e-166;  
 Matches 559; Conservative 0; Mismatches 38; Indels 3; Gaps 1;

QY 1 TTGGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTGGGTAAGCCAGTACAGTG 60  
 DB 4295 TTGGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTGGGTAAGCCAGTACAGTG 4354  
 QY 61 GTTCCCAAGAAAGGTGATGACAGTGTGATGACAGATGAGAGATGATGATGATGATGATGATGATGAT 120  
 DB 4355 GTTCCCAAGAAAGGTGATGACAGTGTGATGACAGATGAGAGATGATGATGATGATGATGATGATGAT 4414  
 QY 121 CGAAGTGTCTGTTGGGATGTGTATGACTATCGCAAGCTGTAAGTAAGCCACACG 180  
 DB 4415 CGAAGTGTCTGTTGGGATGTGTATGACTATCGCAAGCTGTAAGTAAGCCACACG 4474  
 QY 181 AAGGACATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 DB 4475 AAGGACATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4554  
 QY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 DB 4535 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4594  
 QY 301 CAGGGAAGACGGCTTACATGCCCCCTTGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 360  
 DB 4595 CAGGGAAGACGGCTTACATGCCCCCTTGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4654  
 QY 361 GGGTATGTAATGACACACCATTTTCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 420  
 DB 4655 GGGTATGTAATGACACACCATTTTCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 4714  
 QY 421 GTGGAAGAAAGCATGAGTATTTATGACGACTTCTCGGTTTGGACCCCTCAATTTGAC 480



PT New nucleic acid molecules for imparting agronomically significant  
PT characters to plants, especially soybean  
XX  
PS Claim 36(a); Page 113; 118pp; English.

CC The present sequence represents the reverse transcriptase gene of  
CC retrovirus-like element (retroelement) cyclops of pea. The  
CC invention provides molecular tools in the form of retroelements and  
CC retroelement-containing vectors, cells and plants. Methods are  
CC provided for introducing the retroelements into cells, especially  
CC when the retroelement carries at least 1 agronomically-significant  
CC characteristic (ACS). In a preferred method, a helper cell line  
CC which expresses gag, pol and env sequences is used to enable  
CC transfer of a secondary construct which carries an ASC and has  
CC retroelement sequences that allow for replication and integration.  
CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
CC sequence selected from a retroelement primer binding site, envelope,  
CC gag, integrase, reverse transcriptase, protease or RNase-H sequence  
CC (see AA235254-61). Also provided are plant retroviral particles that  
CC comprise a retrovirus protein encoded by a nucleic acid sequence  
CC encoding a plant retroviral envelope protein and a nucleic acid  
CC including e.g. the present sequence. These plant retroviral  
CC particles are used to transfer the nucleic acids into plant cells.

XX Sequence 597 BP; 158 A; 121 C; 152 G; 166 T; 0 other;

Query Match 52.3%; Score 313.8; DB 21; Length 597;  
Best Local Similarity 70.4%; Pred. No. 8.4e-96;  
Matches 420; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 1 TTGGAGCGCTGGGCTCATATATACCCCATCTCTGACAGCGCTGGGTAAGCCAGTACAGTG 60  
DB 1 TTGGATCGAGAAATGATTACCCGATCTCGGATACATCCATGGGTACGTCCTCCGCGATGTG 60  
QY 61 GTTCCAGAGAAAGGTGGAATACAGTGTAGAGATGAGAGAAATGATGATACCAACA 120  
DB 61 GTTCCAGAGAAAGGTGGAATACAGTGTAGAGATGAGAGAAATGATGATACCAACA 120  
QY 121 CGAAGCTGCACTGTTGGCGAATGTGTATCGACTATCGAAGCTGATTAAGCCACAGG 180  
DB 121 AAGGTTCAACGGGTGGAGATGTGTATGAAATAGCGGTGTAATCCGCAACTCGA 180  
QY 181 AAGGACATTTTCCCTTACCTTTCATGATGATCAGATCGTGAAGAGACTTCAGAGGACGCA 240  
DB 181 AAGGACATTTTCCACTCCGTTCAATGATGATCAATGCTGGAAGAACTCTCCGGCAACA 240  
QY 241 TACTACTGTTTCTTGATGATGATCTCGGATTAACAACAGATGCGGTAGACCCAGAGAT 300  
DB 241 TACTATGTTTCTTGATGATGATCTCGGATTAACAACAGATGCGGTAGACCCAGAT 300  
QY 301 CAGAGAGAGAGCGGCTTACATGAGCCCTTGGCGCTTTCCTTACAGAGATGCCATTC 360  
DB 301 CATTAAAGAGCGGCTTTCACATGTCGTTTGGAGTGTTCGATACGGAAGAAATGTCCTTT 360  
QY 361 GGGTATATATGACACGACCATTTAGAGGTGATGTCGTCGCTTTTTCAGACATG 420  
DB 361 GGGTGTGACATGACACGACCATTTCAAGATGTGTCAGACATTTTTCGACACTT 420  
QY 421 GTGAGAGAAAGCATCGAGGTATTATGACACATCTCGGTTTGTGGACCCCTCATTTGAC 480  
DB 421 AATGAGAGAAAGCATCGAGGTATTATGACACATCTCGGTTTGTGGTATTCCTTTAGT 480  
QY 481 AGCTGTTTGGAGACCTAGAGAGGTACTTTCAGAGGTGAGAGAGACTTAATTTGATAGT 540  
DB 481 TTATGCTTGGCAAACTTGAAACGCTGTTGAAAGATGTGTAAAGCCAAATCTTGTCCTT 540  
QY 541 AATGAGAGAAAGTCTCATTTTCTGAGAGGACATGCTTCAGGACCAAGAT 597  
DB 541 AATGAGAGAAAGTCTCATTTTCTGAGAGGACATGCTTCAGGACCAAGAT 597

RESULT 9  
AAF22302/c

ID AAF22302 standard; DNA; 94895 BP.

XX AAF22302;

XX 20-MAR-2001 (first entry)

DE BAC containing repeats from centromeres 1-4 #25.

XX Centromere; microsome; vector; ds.

XX Arabidopsis thaliana.

XX WO20005325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0135584.

XX 17-SEP-1999; 99US-0134603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhagen G, Keith K;

XX WPI; 2000-587529/55.

XX Claim 102; Page 804-819; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited microsome which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;

Query Match 51.3%; Score 307.8; DB 21; Length 94895;

Best Local Similarity 69.6%; Pred. No. 1.5e-92;

Matches 417; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 1 TTGGAGCGCTGGGCTCATATATACCCCATCTCTGACAGCGCTGGGTAAGCCAGTACAGTG 60

DB 35132 TTAGATGTTGGGCTTATCTACCCCTATCTGTGATGACAGTGGTAAATCTCCAGTTTCG 35073

QY 61 GTTCCAGAGAAAGGTGGAATACAGTGTAGAGATGAGAGAAATGATGATACCAACA 120

DB 35072 GTTCCAGAGAAAGGTGGAATACAGTGTAGAGATGAGAGAAATGATGATACCAACA 35013

QY 121 CGAAGCTGCACTGTTGGCGAATGTGTATCGACTATCGAAGCTGATTAAGCCACAGG 180

DB 35012 AGGACTATTAAGTGCATAGATGTGTATGACTATAGAGAAATGATGATGCTCTAGA 34953

QY 181 AAGGACATTTTCCCTTACCTTTCATGATGATCAGATCGTGAAGAGACTTCAGAGGACGCA 240

DB 34952 AAGGACATTTTCCCTTACCTTTCATGATGATCAGATCGTGAAGAGACTTCAGAGGACGCA 34893

QY 241 TACTACTGTTTCTTGATGATGATCTCGGATTAACAACAGATGCGGTAGACCCAGAGAT 300

DB 34892 TACTATGCTTCTTGATGATGATCTCGGATTAACAACAGATGCGGTAGACCCAGAGAT 34833

QY 301 CAGAGAGAGAGCGGCTTTCATATGCCCCCTTGGCGCTTTCCTTACAGAGAGATGCCATTC 360

DB 34832 CAGAGAGAGAGCGGCTTTCATATGCCCCCTTGGCGCTTTCCTTACAGAGAGATGCCATTC 34773

```
OY 361 GGGTATGTATGACACACCATTTTCAGAGGTGATGCTGGCCATTTTTCAGACATG 420
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34772 GGTCTATGTAATGACACCATTTCTTTCACGGCTGTATGACTTCTATTTCTCGGATTTG 34713
OY 421 GTGGAGAAAAGCATGAGGTATTTATGACGACTTCGCTTTTGGACCCATTTTAC 480
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34712 ATAGAGGAGATGTAAGAGTATTCATGATGATTTTCTGTATGCTGCTCTCTCTC 34653
OY 481 AGCTGTTTGGAGAGGAGGTATTCAGAGGTGCGAAGAGACTTACTGTACTG 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34652 TCGTCTTTTCTTGAACCTGTGTAGGGTACTTAAAGATGTGAAGACAAACATGCTG 34593
OY 541 AATTGGAAAAGTGTCTATTTCATGTTTCGAGAGGCGATAGTCTTACGCCACAAGATCTC 599
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34592 AACTGGAGAAATGCCATTTTCATGTTTGAAGAAGCATGCTTTTGGGCCACAAATTTTC 34534

RESULT 10
AAF22305
ID AAF22305 standard; DNA: 1082138 BP.
AC AAF22305;
XX
XX 20-MAR-2001 (first entry)
DT
XX
DE Arabidopsis thaliana chromosome 4 centromere.
XX
XX Centromere; mitochondrion; vector; ds.
XX
XX Arabidopsis thaliana.
OS
XX
XX WO200055325-A2.
XX
XX 21-SEP-2000.
PD
XX
XX 17-MAR-2000; 2000WO-US07392.
PF
XX
XX 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0135384.
PR 17-SEP-1999; 99US-0154603.
XX
XX (UYCH-) UNIV CHICAGO.
PA
XX
XX Preuss D, Copenhaver G, Kelth K;
XX
XX WPI: 2000-587529/55.
DR
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
XX
XX producing stably inherited mitochondria which can serve as vectors for
XX
XX the construction of transgenic plant and animal cells
XX
XX Claim 68; Page 977-1388; 1449Pp; English.
PS
XX
XX The present invention relates to a recombinant DNA construct of a plant
XX
XX (Arabidopsis thaliana) centromere. The constructs are useful for
XX
XX producing stably inherited mitochondria which can serve as vectors for
XX
XX the construction of transgenic plant and animal cells expressing
XX
XX selected proteins such as hormones, enzymes, interleukins, clotting
XX
XX factors, cytokines, antibodies, and growth factors.
XX
XX Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;
SQ

Query Match 51.38; Score 307.8; DB 21; Length 1082138;
Best Local Similarity 69.68; Pred. No. 5.9e-92;
Matches 417; Conservative 0; Mismatches 182; Indels 0; Gaps 0;
```

```
OY 1 TTGAGAGCTGGCTCATATACCCCATCTGACAGGCTGGTGAAGCCAGTACAGTGC 60
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1047007 TTAGATGTTGGGTTATCTACCTATCTCTGATAGACATGGGTATCTCCAGTTCAATTGC 1047066
OY 61 GTTCCCAAGAAAGTGAATGACAGTGTAGAGATGAGAGATGACTTGATACCAACA 120

RESULT 11
AA35264
ID AA35264 standard; DNA: 600 BP.
AC AA35264;
XX
XX 27-MAR-2000 (first entry)
DT
XX
XX Arabidopsis retroelement athila reverse transcriptase gene.
XX
XX Retroelement; retrovirus; transgenic plant; gene transfer;
XX
XX reverse transcriptase gene; athila; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX WO9960842-A2.
XX
XX 02-DEC-1999.
PD
XX
XX 28-MAY-1999; 99WO-US11858.
PF
XX
XX 29-MAY-1998; 98US-0087125.
PR 28-MAY-1999; 99US-0087125.
PR
XX
XX (WRIG/) WRIGHT D A.
XX
XX (VOIT/) VOITAS D F.
XX
XX Wright DA, Voytas DF;
XX
XX WPI: 2000-105586/09.
DR
XX
XX New nucleic acid molecules for imparting agronomically significant
XX
XX characters to plants, especially soybean
XX
XX Claim 36(a); Page 112-113; 118Pp; English.
PS
XX
XX The present sequence represents the reverse transcriptase gene of
```













GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:44:01 ; Search time 72.6415 Seconds  
(without alignments)  
2535.070 Million cell updates/sec

Title: US-09-965-553-11

Perfect score: 600

Sequence: 1 tttagagctggcctcatata.....tcttagccacaagatctca 600

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing First 45 summaries

Database :

Issued Patents\_NA: \*  
1: /cgn2\_6/prodata/1/lna/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/lna/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/lna/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/lna/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/lna/PCrUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/lna/Backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	600	100.0	600	US-09-322-478-11	Sequence 11, Appl
2	600	100.0	12286	US-09-322-478-17	Sequence 17, Appl
3	571.2	95.2	4609	US-09-322-478-21	Sequence 21, Appl
4	568.2	94.7	597	US-09-322-478-34	Sequence 34, Appl
5	532	88.7	9829	US-09-322-478-19	Sequence 19, Appl
6	523.2	87.2	12571	US-09-322-478-20	Sequence 20, Appl
7	345.6	57.6	9139	US-09-322-478-22	Sequence 22, Appl
8	313.8	52.3	597	US-09-322-478-29	Sequence 29, Appl
9	304.6	50.8	600	US-09-322-478-27	Sequence 27, Appl
10	172.4	28.7	10482	US-09-322-478-23	Sequence 23, Appl
11	69	11.5	1347	US-08-855-449-2	Sequence 2, Appl
12	55.4	9.2	7568	US-08-694-869-2	Sequence 2, Appl
13	55.4	9.2	7568	US-09-349-546-2	Sequence 2, Appl
14	48.6	8.1	31491	US-09-360-186-1	Sequence 1, Appl
15	33.8	5.6	1421	US-09-592-054-5	Sequence 5, Appl
16	33.8	5.6	1569	US-09-592-054-3	Sequence 3, Appl
17	33	5.5	360	US-09-605-785-438	Sequence 438, App
18	33	5.5	360	US-09-439-313-438	Sequence 438, App
19	33	5.5	360	US-09-352-616A-438	Sequence 438, App
20	32.2	5.4	4127	US-09-592-054-7	Sequence 7, Appl
21	31.4	5.2	7218	US-08-232-463-14	Sequence 14, Appl
22	31.4	5.2	1977	US-09-231-529-2	Sequence 2, Appl
23	31	5.2	1977	US-08-977-816-2	Sequence 2, Appl
24	30.6	5.1	4199	US-09-204-117B-1	Sequence 1, Appl
25	30.2	5.0	1734	US-08-663-808-1	Sequence 1, Appl
26	30.2	5.0	1734	US-09-332-740-1	Sequence 1, Appl
27	30.2	5.0	1734	US-09-188-496-1	Sequence 1, Appl

28	30.2	5.0	1734	US-09-368-282-1	Sequence 1, Appl
29	30.2	5.0	1734	US-09-566-708A-1	Sequence 1, Appl
30	29.6	5.0	491	US-09-404-879A-33	Sequence 33, Appl
31	29.6	4.9	2327	US-08-835-170-3	Sequence 3, Appl
32	29.6	4.9	2327	US-09-359-257-3	Sequence 3, Appl
33	29.6	4.9	2327	US-09-371-674-3	Sequence 3, Appl
34	29.4	4.9	1294	US-09-425-578-1	Sequence 1, Appl
35	29.4	4.9	1444	US-08-336-427-9	Sequence 9, Appl
36	29.4	4.9	1447	US-08-236-427-2	Sequence 2, Appl
37	29.4	4.9	5197	US-08-131-365B-53	Sequence 53, Appl
38	29.4	4.9	5197	US-08-668-123-53	Sequence 53, Appl
39	29	4.8	2919	US-08-434-000A-3	Sequence 3, Appl
40	29	4.8	2919	US-09-312-157-3	Sequence 3, Appl
41	28.8	4.8	1314	US-08-800-682-1	Sequence 1, Appl
42	28.8	4.8	4403765	US-09-103-840A-2	Sequence 2, Appl
43	28.8	4.8	4411529	US-09-103-840A-1	Sequence 1, Appl
44	28.6	4.8	1622	US-09-232-191-12	Sequence 12, Appl
45	28.6	4.8	1622	US-09-232-200-12	Sequence 12, Appl

#### ALIGNMENTS

RESULT 1					
US-09-322-478-11					
Sequence 11, Application US/09322478					
Patent No. 6331662					
GENERAL INFORMATION:					
APPLICANT: Wright, David A.					
FILE OF INVENTION: Plant Retroelements and Methods Related Thereto					
TITLE REFERENCE: P-1065 ISURF Plant Retroelement					
CURRENT APPLICATION NUMBER: US/09/322,478					
EARLIER FILING DATE: 1999-05-28					
EARLIER APPLICATION NUMBER: 60/087125					
NUMBER OF SEQ ID NOS: 41					
SOFTWARE: Patentln Ver. 2.0					
SEQ ID NO 11					
LENGTH: 600					
TYPE: DNA					
ORGANISM: Artificial Sequence					
FEATURE:					
OTHER INFORMATION: Description of Artificial Sequence: plant					
US-09-322-478-11					
Query Match					
Best Local Similarity 100.0%; Pred. No. 6.4e-199;					
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG	60		
DB	1	TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG	60		
QY	61	GTTCCTCAAAAGAGTGAATGACAGTGTACAGATGAGAGATGACTGATACACACA	120		
DB	61	GTTCCTCAAAAGAGTGAATGACAGTGTACAGATGAGAGATGACTGATACACACA	120		
QY	121	CGAAGCTACAGTGGTGGGATGTATGATGATGCAAGCTGATGAAGCCACACCG	180		
DB	121	CGAAGCTACAGTGGTGGGATGTATGATGATGCAAGCTGATGAAGCCACACCG	180		
QY	181	AAGGACCATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGAT	240		
DB	181	AAGGACCATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGAT	240		
QY	241	TACTACTGTTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGAT	300		
DB	241	TACTACTGTTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGAT	300		
QY	301	CAGAGAGACGCGCTTACATGCGCTTGGGCTTCTTCTTACAGAGAGATCCATTC	360		
DB	301	CAGAGAGACGCGCTTACATGCGCTTGGGCTTCTTCTTACAGAGAGATCCATTC	360		

Db 301 CAGGAGAAAGCGCTTTACATGCCCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 360  
QY 361 GGGTTATGTAATGACACCGACCATTTTACAGAGTGCATGCTGGCCATTTTTCAGACATG 420  
Db 361 GGGTTATGTAATGACACCGACCATTTTACAGAGTGCATGCTGGCCATTTTTCAGACATG 420  
QY 421 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTGC 480  
Db 421 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTGC 480  
QY 481 AGCTGTTGAGGAACCTAGAGAGGTACTTTCAGAGTGGGAGAGACTTAATCTGACTG 540  
Db 481 AGCTGTTGAGGAACCTAGAGAGGTACTTTCAGAGTGGGAGAGACTTAATCTGACTG 540  
QY 541 AATTGGAAAAAGTGTCAATTTTCATGTTGTCGAGAGGCGATAGTCTTAGGCCACAAGATCTCA 600  
Db 541 AATTGGAAAAAGTGTCAATTTTCATGTTGTCGAGAGGCGATAGTCTTAGGCCACAAGATCTCA 600

RESULT 2  
US-09-322-478-17  
; Sequence 17, Application US/09322478  
; Patent No. 6331662  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/322.478  
; EARLIER FILING DATE: 1998-05-28  
; EARLIER APPLICATION NUMBER: 60/087125  
; EARLIER FILING DATE: 1998-05-29  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 12286  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: plant  
; OTHER INFORMATION: retroelement sequence  
US-09-322-478-17

Query Match 100.0%; Score 600; DB 4; Length 12286;  
Best Local Similarity 100.0%; Pred. No. 3.8e-198;  
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGGAGGCTGGGCTCATATACCCCATCTCTGACAGCCGTTGGGTAAGCCCACTACAGTG 60  
Db 4278 TTGGAGGCTGGGCTCATATACCCCATCTCTGACAGCCGTTGGGTAAGCCCACTACAGTG 4337  
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGGTAGAGATGAGAGAAATGACTGTATACCAACA 120  
Db 4338 GTTCCCAAGAAAGGTGGAATGACAGTGGTAGAGATGAGAGAAATGACTGTATACCAACA 4397  
QY 121 CGAAGTGTCACTGGTGGCGAATGTATGCACTATGCGAAGCTGAATGAACCAACACGG 180  
Db 4398 CGAAGTGTCACTGGTGGCGAATGTATGCACTATGCGAAGCTGAATGAACCAACACGG 4457  
QY 181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 240  
Db 4458 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 4517  
QY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
Db 4518 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4577  
QY 301 CAGGAGAAAGCGGCTTTTACATGCGCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 360  
Db 4578 CAGGAGAAAGCGGCTTTTACATGCGCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 4637  
QY 361 GGGTTATGTAATGACACCGACCATTTTTCAGAGTGCATGCTGGCCATTTTTCAGACATG 420

Db 4638 GGGTTATGTAATGACACCGACCCATTTTCAGAGTGCATGCTGGCCATTTTTCAGACATG 4697  
QY 421 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTGC 480  
Db 4698 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTGC 4757  
QY 481 AGCTGTTGAGGAACCTAGAGAGGTACTTTCAGAGTGGGAGAGACTTAATCTGACTG 540  
Db 4758 AGCTGTTGAGGAACCTAGAGAGGTACTTTCAGAGTGGGAGAGACTTAATCTGACTG 4817  
QY 541 AATTGGAAAAAGTGTCAATTTTCATGTTGTCGAGAGGCGATAGTCTTAGGCCACAAGATCTCA 600  
Db 4818 AATTGGAAAAAGTGTCAATTTTCATGTTGTCGAGAGGCGATAGTCTTAGGCCACAAGATCTCA 4877

RESULT 3  
US-09-322-478-21  
; Sequence 21, Application US/09322478  
; Patent No. 6331662  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/322.478  
; EARLIER FILING DATE: 1998-05-28  
; EARLIER APPLICATION NUMBER: 60/087125  
; EARLIER FILING DATE: 1998-05-29  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 21  
; LENGTH: 4609  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-322-478-21

Query Match 95.2%; Score 571.2; DB 4; Length 4609;  
Best Local Similarity 97.0%; Pred. No. 2.2e-188;  
Matches 582; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 TTGGAGGCTGGGCTCATATACCCCATCTCTGACAGCCGTTGGGTAAGCCCACTACAGTG 60  
Db 1174 TTGGAGGCTGGGCTCATATACCCCATCTCTGACAGCCGTTGGGTAAGCCCACTACAGTG 1233  
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGGTAGAGATGAGAGAAATGACTGTATACCAACA 120  
Db 1234 GTTCCCAAGAAAGGTGGAATGACAGTGGTAGAGATGAGAGAAATGACTGTATACCAACA 1293  
QY 121 CGAAGTGTCACTGGTGGCGAATGTATGCACTATGCGAAGCTGAATGAACCAACACGG 180  
Db 1294 CGAAGTGTCACTGGTGGCGAATGTATGCACTATGCGAAGCTGAATGAACCAACACGG 1353  
QY 181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 240  
Db 1354 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 1413  
QY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
Db 1414 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1473  
QY 301 CAGGAGAAAGCGGCTTTTACATGCGCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 360  
Db 1474 CAGGAGAAAGCGGCTTTTACATGCGCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 1533  
QY 361 GGGTTATGTAATGACACCGACCATTTTTCAGAGTGCATGCTGGCCATTTTTCAGACATG 420  
Db 1534 GGGTTATGTAATGACACCGACCATTTTTCAGAGTGCATGCTGGCCATTTTTCAGACATG 1593  
QY 421 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTGC 480  
Db 1594 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTGC 1653  
QY 481 AGCTGTTGAGGAACCTAGAGAGGTACTTTCAGAGTGGGAGAGACTTAATCTGACTG 540

Db 1624 AGCTGTTGAGAACCTAGAAATGTTACTTCAGAGAGTGGCTAGAGACTAATGTTGTTACTG 1713  
QY 541 AATTGGGAAAAGTGCATTTCATGTTGTCGAGAGGGGATAGTCTAGGCCCAAGATCTCA 600  
Db 1714 AATTGGGAAAAGTGCATTTCATGTTGTTGTCGAGAGGGGATAGTCTAGGCCCAAGATCTCA 1773

## RESULT 4

US-09-322-478-34  
; Sequence 34, Application US/09322478  
; Patent No. 633162  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/322,478  
; EARLIER FILING DATE: 1999-05-28  
; EARLIER APPLICATION NUMBER: 60/087125  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 34  
; LENGTH: 597  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-322-478-34

Query Match 94.7%; Score 568.2; DB 4; Length 597;  
Best Local Similarity 97.0%; Pred. No. 7.5e-188;  
Matches 579; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 TTGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG 60  
Db 1 TTGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG 60  
QY 61 GTTCCCAAGAAAGTGGATGACAGTGTGATGACATGACAGTGTGATGACAGTGTGATGACAG 120  
Db 61 GTTCCCAAGAAAGTGGATGACAGTGTGATGACATGACAGTGTGATGACAGTGTGATGACAG 120  
QY 121 CGAAGTGTCTACTGTTGGCGAATGTGTATGACATGACAGTGTGATGACAGTGTGATGACAG 180  
Db 121 CGAAGTGTCTACTGTTGGCGAATGTGTATGACATGACAGTGTGATGACAGTGTGATGACAG 180  
QY 181 AAGGACCTTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
Db 181 AAGGACCTTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
QY 241 TACTACTGTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
Db 241 TACTACTGTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
QY 301 CAGGAGAGAGCGGCTTTACATGCCCTTTGGCGTCTTGGCTTACAGAAAGATGCCATTC 360  
Db 301 CAGGAGAGAGCGGCTTTACATGCCCTTTGGCGTCTTGGCTTACAGAAAGATGCCATTC 360  
QY 361 GGGTATATATGACACGACCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATGAT 420  
Db 361 GGGTATATATGACACGACCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATGAT 420  
QY 421 GTGGAAGAAAGCATGAGATATTTATGACAGCTTCTGCTTTTGGAGCCCATTTTGGAC 480  
Db 421 GTGGAAGAAAGCATGAGATATTTATGACAGCTTCTGCTTTTGGAGCCCATTTTGGAC 480  
QY 481 AGCTGTTTGGAGAACCTAGAGAGGTTACTTACAGAGTGTGAGAGACTAATCTTGTACTG 540  
Db 481 AGCTGTTTGGAGAACCTAGAGAGGTTACTTACAGAGTGTGAGAGACTAATCTTGTACTG 540  
QY 541 AATTGGGAAAAGTGCATTTCATGTTGTCGAGAGGGGATAGTCTAGGCCCAAGATCTCA 597  
Db 541 AATTGGGAAAAGTGCATTTCATGTTGTCGAGAGGGGATAGTCTAGGCCCAAGATCTCA 597

## RESULT 5

US-09-322-478-19  
; Sequence 19, Application US/09322478  
; Patent No. 633162  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/322,478  
; EARLIER FILING DATE: 1999-05-28  
; EARLIER APPLICATION NUMBER: 60/087125  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 9829  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-322-478-19

Query Match 88.7%; Score 532; DB 4; Length 9829;  
Best Local Similarity 94.0%; Pred. No. 1.5e-174;  
Matches 564; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

QY 1 TTGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG 60  
Db 4271 TTGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG 4330  
QY 61 GTTCCCAAGAAAGTGGATGACAGTGTGATGACATGACAGTGTGATGACAGTGTGATGACAG 120  
Db 4331 GTTCCCAAGAAAGTGGATGACAGTGTGATGACATGACAGTGTGATGACAGTGTGATGACAG 4390  
QY 121 CGAAGTGTCTACTGTTGGCGAATGTGTATGACATGACAGTGTGATGACAGTGTGATGACAG 180  
Db 4391 CGAAGTGTCTACTGTTGGCGAATGTGTATGACATGACAGTGTGATGACAGTGTGATGACAG 4450  
QY 181 AAGGACCTTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
Db 4451 AAGGACCTTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
QY 241 TACTACTGTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
Db 4511 TACTACTGCTT-TTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4569  
QY 301 CAGGAGAGAGCGGCTTTACATGCCCTTTGGCGTCTTGGCTTACAGAAAGATGCCATTC 360  
Db 4570 CAGGAGAGAGCGGCTTTACATGCCCTTTGGCGTCTTGGCTTACAGAAAGATGCCATTC 4629  
QY 361 GGGTATATATGACACGACCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATGAT 420  
Db 4630 GGGTATATGACACGACCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATGAT 4689  
QY 421 GTGGAAGAAAGCATGAGATATTTATGACAGCTTCTGCTTTTGGAGCCCATTTTGGAC 480  
Db 4690 GTGGAAGAAAGCATGAGATATTTATGACAGCTTCTGCTTTTGGAGCCCATTTTGGAC 4749  
QY 481 AGCTGTTTGGAGAACCTAGAGAGGTTACTTACAGAGTGTGAGAGACTAATCTTGTACTG 540  
Db 4750 AACTATTTGAGAGAACCTAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 4809  
QY 541 AATTGGGAAAAGTGCATTTCATGTTGTCGAGAGGGGATAGTCTAGGCCCAAGATCTCA 600  
Db 4810 AATTGGGAAAAGTGCATTTCATGTTGTCGAGAGGGGATAGTCTAGGCCCAAGATCTCA 4869

## RESULT 6

US-09-322-478-20  
; Sequence 20, Application US/09322478  
; Patent No. 633162  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.

```

; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; CURRENT FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 12571
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-20

Query Match      87.2%; Score 523.2; DB 4; Length 12571;
Best Local Similarity 93.2%; Pred.No. 2,1e-171;
Matches 559; Conservative 0; Mismatches 38; Indels 3; Gaps 1;

OY 1 TTGGAGGCTGGGCTCATATACCCCTCTCTGACAGCGCTGGGTAAGCCAGTACAGTG 60
DB 4295 TTGGAGGCTGGGCTCATATACCCCTCTCTGACAGCGCTGGGTAAGCCAGTACAGTG 4354
OY 61 GTTCCCAAGAAAGGTGATGACAGTGTGATGAGAGGATGACTTGATACCAACA 120
DB 4355 GTTCCCAAGAAAGGTGATGACAGTGTGATGAGAGGATGACTTGATACCAACA 4414
OY 121 CGAAGCTGACGTGGTGGCAATGTATGACATGTCGCAAGCTGCAATGCAAGCCACAGG 180
DB 4415 CGAAGCTGACGTGGTGGCAATGTATGACATGTCGCAAGCTGCAATGCAAGCCACAGG 4474
OY 181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 240
DB 4475 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 4534
OY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 4535 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4594
OY 301 CAGGAGAAGAGGCGCTTTTACATGCGCCCTTGGCGCTTCTTGTACAGAGATGCAATTC 360
DB 4595 CAGGAGAAGAGGCGCTTTTACATGCGCCCTTGGCGCTTCTTGTACAGAGATGCAATTC 4654
OY 361 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 4655 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4714
OY 421 GTGAGAGAAAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
DB 4715 GTGAGAGAAAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4774
OY 481 AGCTGTTTGAAGAACTGAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 540
DB 4775 A--GTTTGAAGAACTGAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 4831
OY 541 AATTGGGAAAAGTGTCTATTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 600
DB 4832 AACTGGGAAAAGTGTCTATTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 4891

RESULT 7
US-09-322-478-22
; Sequence 22, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; CURRENT FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
;

```

```

; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 9139
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-22

Query Match      57.6%; Score 345.6; DB 4; Length 9139;
Best Local Similarity 73.5%; Pred.No. 1,1e-109;
Matches 441; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

OY 1 TTGGAGGCTGGGCTCATATACCCCTCTCTGACAGCGCTGGGTAAGCCAGTACAGTG 60
DB 1646 TAGAAGCAGGCGCTTATTTTACCCCTCTCTGATGATGATGATGATGATGATGATGATG 1705
OY 61 GTTCCCAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
DB 1706 GTTCCCAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1765
OY 121 CGAAGCTGACGTGGTGGCAATGTATGACATGTCGCAAGCTGCAATGCAAGCCACAGG 180
DB 1766 AGGACGTGACCGGGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1825
OY 181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 240
DB 1826 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 1885
OY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 1886 TATTATGTTTCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1945
OY 301 CAGGAGAAGAGGCGCTTTTACATGCGCCCTTGGCGCTTCTTGTACAGAGATGCAATTC 360
DB 1946 CAGGAGAAGAGGCGCTTTTACATGCGCCCTTGGCGCTTCTTGTACAGAGATGCAATTC 2005
OY 361 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 2006 GGTCTGGAATGCGCCGCTTACATTCAGAGGATGATGATGATGATGATGATGATGATG 2065
OY 421 GTGAGAGAAAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
DB 2066 GTGAGAGAAAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2125
OY 481 AGCTGTTTGAAGAACTGAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 540
DB 2126 GGTGCTGCTTATTAATCTTGAAGAGATTTACAGAGATGTAAGAGTCCAACTAGTTCTC 2185
OY 541 AATTGGGAAAAGTGTCTATTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 600
DB 2186 AATTGGGAAAAGTGTCTATTTTCATGATGATGATGATGATGATGATGATGATGATGATG 2245

RESULT 8
US-09-322-478-29
; Sequence 29, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; CURRENT FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
;
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Pisum sativum
US-09-322-478-29

```



Db 4124 GAATGTCATGATATCGAAGTTGAATTAATGCCACTTGAAAGCAATTAATCCACTCC 4183  
OY 200 CTTTCATGATGATGCTGAGAGACTTGCAGGCGAGCATAGTACTTTCTTGAGG 259  
Db 4184 CTTTCATGAGACCATGCTTGTGAGAGCTGCGAAGCAATTAATTTCTTGAGG 4243  
OY 260 GATACCTGGGATACACCAAGATCCGGGTAGACCCAGAGATCAGA 305  
Db 4244 GATATCTAGTTACAAATAGATTGCTATAGATATCAAGATCAAGA 4289

RESULT 11  
US-08-855-449-2/c  
Sequence 2, Application US/08855449  
Patent No. 5910412

GENERAL INFORMATION:  
APPLICANT: AKAMATSU, TOYOKAZU  
TITLE OF INVENTION: METHOD FOR IDENTIFYING THE SEX OF  
TITLE OF INVENTION: SPINACH BY DNA MARKERS  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESS: P.C. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/855,449  
FILING DATE: 13-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 119124/1996  
FILING DATE: 14-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 7828-0003-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ. ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1347 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: SPINACH  
US-08-855-449-2

Query Match 11.5%; Score 69; DB 2; Length 1347;  
Best Local Similarity 51.1%; Pred. No. 6.6e-14;  
Matches 162; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

OY 139 GCAATGCTATGACATGCAAGCTGAATGAAGCCACGAGACATTTCCCTTA 198  
Db 1111 CAAATGTCGCTGATTAAGATATCTTAACAGACTAGCCCTAAAGACGATTTCCACTG 1052  
OY 199 CTTTCATGATGATGCTGAGAGACTTGCAGGCGAGCATAGTACTTTCTTGAGT 258  
Db 1051 CCACTCAACGACATTTCTGTGACAAACACCAATTAATGCTTCTTTTATGAGAC 992  
OY 259 GGATACCTGGGATACACCAAGATCGGGTAAAGACCCAGAGATCAGAGAAAGACGGCTTT 318

Db 991 TGTAGCAGAGCTACCAATTCGATTCCTCATACAGAGACGACATGAGAAACAACTTC 932  
OY 319 ACATGCCCTTTGGCGCTTTGCTTACAGAGATGCCATTCGGGTTATGTAATGCACCA 378  
Db 931 ATCACTAAGTGGGGTACATATTCCTAAACAGTTATGCCGTTTGACACGAAACACGGGG 872  
OY 379 GCCACATTTACAGAGTGCATGCTGCCATTTTTCAGACATGTGGAGAAAGCATCGAG 438  
Db 871 GCTACCTATCAATTAACACCAACCAACCATCATTAAGCAGCATGATTAAACAGAAATTGAG 812  
OY 439 GTATTTATGAGCAGCTT 455  
Db 811 GAATATGTCGACGACAT 795

RESULT 12  
US-08-694-869-2  
Sequence 2, Application US/08694869  
Patent No. 5994123

GENERAL INFORMATION:  
APPLICANT: Olaszewski, N.  
APPLICANT: Tzaifir, I.  
APPLICANT: Somers, D. A.  
APPLICANT: Lockhart, B.  
APPLICANT: Torbert, K.  
TITLE OF INVENTION: SUGARCANE BACILLIFORM VIRUS  
TITLE OF INVENTION: PROMOTER  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.  
STREET: P.O. Box 2938  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/694,869  
FILING DATE: 09-AUG-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Woessner, Warren D  
REGISTRATION NUMBER: 30,440  
REFERENCE/DOCKET NUMBER: 600.369051  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-373-6900  
TELEFAX: 612-339-3061  
TELEX:  
INFORMATION FOR SEQ. ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7568 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-694-869-2

Query Match 9.2%; Score 55.4; DB 2; Length 7568;  
Best Local Similarity 46.6%; Pred. No. 9.8e-09;  
Matches 213; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

OY 140 GAATGCTATGACATGCAAGCTGAATGAAGCCACGAGACATTTCCCTTAC 199  
Db 5482 GACTGCTTTTCAATTAACAGAGGCTAAATGACAAACATAGCGCGGATCAATTAATTCATTGC 5541



QY 200 CTTTCATGATCAGATGCTGGAGAGACTTGCAGGCGAGCATACACTGTTCTTGATG 259  
 Db 5542 CCGGATCATATGCTCTACTATAAAATGTTGCAAGACAAAGATCTCTCAAGTTTGAT 5601  
 QY 260 GATCTCGGGATACACACAGATCGGGTAGACCCAGAGATCAGAGAGAGCGCTTTA 319  
 Db 5602 TGAAGAGCGGGTTTCATCAAGTCGCCATGATGAGAAAGTATTCATTAACAGCATTTT 5661  
 QY 320 CATGCCCTTTGGGCTTTTGGTTTACAGAGATGCCATTTGGGTTATTAATGACACAG 379  
 Db 5662 CAGATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5721  
 QY 380 CCACATTTAGAGAGTGCATGCTGGCCATTTTTCAGACATGCTGAGAAAGAGATGAGG 439  
 Db 5722 CAATCTTCAGAGAAAGAAATGAGCCAGTGTTCAGAGAAACAGAGG--TTCATAGCTG 5778  
 QY 440 TATTATGAGAGACTTCTCGGTTTGGACCTCATTTGACAGCTTTTGAGAGACCTAG 499  
 Db 5779 TGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5838  
 QY 500 AGAGGCTACTTCAGAGGTGCGAAGAGACTAACTTGTTACTGATTTGGAAAGTGTGAT 559  
 Db 5839 GGAAGATGCTTCAATCTGCAAAAGAGATGACTATCTTGATGCTCGTCAAGTACAGA 5898  
 QY 560 TCATGTTTCAGAGGCGCATGCTCTAGGCCACACAGAT 596  
 Db 5899 TAGAGTTAAGAGGTGATGATTTCTTGGAGACACAAAT 5935

RESULT 13  
 US-09-349-546-2  
 : Sequence 2, Application US/09349546  
 : Patent No. 6093569  
 : GENERAL INFORMATION:  
 : APPLICANT: Olszewski, N.  
 : APPLICANT: Tzafir, I.  
 : APPLICANT: Somers, D.A.  
 : APPLICANT: Lockhart, B.  
 : APPLICANT: Torbert, K.  
 : TITLE OF INVENTION: Sugarcane bacilliform virus promoter  
 : FILE REFERENCE: 600.369US2  
 : CURRENT APPLICATION NUMBER: US/09/349,546  
 : EARLIER FILING DATE: 1999-07-08  
 : EARLIER APPLICATION NUMBER: US 08/694,869  
 : EARLIER FILING DATE: 1996-08-09  
 : EARLIER APPLICATION NUMBER: PCT/IB97/01338  
 : EARLIER FILING DATE: 1997-08-13  
 : NUMBER OF SEQ ID NOS: 10  
 : SOFTWARE: FastSeq for Windows Version 4.0  
 : SEQ ID NO 2  
 : LENGTH: 7568  
 : TYPE: DNA  
 : ORGANISM: sugarcane bacilliform virus  
 : US-09-349-546-2

Query Match 9.2%; Score 55.4; DB 3; Length 7568;  
 Best Local Similarity 46.6%; Pred. No. 9, 8e-09;  
 Matches 213; Conservative 0; Mismatches 241; Indels 3; Gaps 1;  
 QY 140 GAATGTATGACTATTCGCAAGCTGAATGAAGCCACAGAGACCAATTCCTCCCTTAC 199  
 Db 5482 GACTGGTTTCAATTAACAGAGGCTAATGACACACAGATGCGGATCAATATTCATGTC 5541  
 QY 200 CTTTCATGATCAGATGCTGGAGAGACTTGCAGGCGAGCATACACTGTTCTTGATG 259  
 Db 5542 CCGGATCATATGCTCTACTATAAAATGTTGCAAGACAAAGATCTCTCAAGTTTGAT 5601  
 QY 260 GATCTCGGGATACACACAGATCGGGTAGACCCAGAGATCAGAGAGAGCGCTTTA 319  
 Db 5602 TGAAGAGCGGGTTTCATCAAGTCGCCATGATGAGAAAGTATTCATTAACAGCATTTT 5661  
 QY 320 CATGCCCTTTGGGCTTTTGGTTTACAGAGATGCCATTTGGGTTATTAATGACACAG 379

Db 5662 CAGATACATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 5721  
 QY 380 CCACATTTACAGAGTGCATGCTGGCCATTTTTCAGACATGCTGAGAAAGATGAGG 439  
 Db 5722 CAATCTTCAGAGAAAGAAATGAGCCAGTGTTCAGAGAAACAGAGG--TTCATAGCTG 5778  
 QY 440 TATTATGAGAGACTTCTCGGTTTGGACCTCATTTGACAGCTTTTGAGAGACCTAG 499  
 Db 5779 TGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5838  
 QY 500 AGAGGCTACTTCAGAGGTGCGAAGAGACTAACTTGTTACTGATTTGGAAAGTGTGAT 559  
 Db 5839 GGAAGATGCTTCAATCTGCAAAAGAGATGACTATCTTGATGCTCGTCAAGTACAGA 5898  
 QY 560 TCATGTTTCAGAGGCGCATGCTCTAGGCCACACAGAT 596  
 Db 5899 TAGAGTTAAGAGGTGATGATTTCTTGGAGACACAAAT 5935

RESULT 14  
 US-09-360-186-1/c  
 : Sequence 1, Application US/09360186  
 : Patent No. 6262343  
 : GENERAL INFORMATION:  
 : APPLICANT: Staskawicz, et al.  
 : TITLE OF INVENTION: Bsz Resistance Gene  
 : FILE REFERENCE: 50687  
 : CURRENT APPLICATION NUMBER: US/09/360,186  
 : EARLIER FILING DATE: 1999-07-23  
 : EARLIER APPLICATION NUMBER: 60/093,957  
 : EARLIER FILING DATE: 1998-07-23  
 : NUMBER OF SEQ ID NOS: 9  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 1  
 : LENGTH: 31491  
 : TYPE: DNA  
 : ORGANISM: Capsicum annuum  
 : US-09-360-186-1

Query Match 8.1%; Score 48.6; DB 4; Length 31491;  
 Best Local Similarity 46.8%; Pred. No. 5, 3e-06;  
 Matches 153; Conservative 0; Mismatches 174; Indels 0; Gaps 0;  
 QY 139 CGAATGTATGACTATTCGCAAGCTGAATGAAGCCACAGAGACCAATTCCTCCCTTA 198  
 Db 24307 CCGATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 24248  
 QY 199 CTTTCATGATCAGATGCTGGAGAGACTTGCAGGCGAGCATACACTGTTCTTGAT 258  
 Db 24247 CCTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 24188  
 QY 259 GGATACCTGGGATACACACAGATCGCGTAGACCCACAGAGATCAGAGAGAGCGCTTT 318  
 Db 24187 CTGCTGCGGTTTACATCAAGTTGAAATTAAGAGATCAGATACCAAGAGAGCGCTTC 24128  
 QY 319 ACATGCCCTTTGGGCTTTTGGTTTACAGAGATGCCATTTGGGTTATTAATGACACA 378  
 Db 24127 CGAACCCATATGTCATCAATTTTATGATGATGATGATGATGATGATGATGATG 24068  
 QY 379 GCCACATTTTCAGAGGTGATGCTGGCCATTTTTCAGACATGATGATGATGATGATG 438  
 Db 24067 GCAGCGTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 24008  
 QY 439 GTATTATGAGAGACTTCTCGGTTT 465  
 Db 24007 GTATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 23981  
 RESULT 15  
 US-09-392-054-5  
 : Sequence 5, Application US/09592054  
 : Patent No. 6440684  
 : GENERAL INFORMATION:

APPLICANT: Berand, Christophe  
APPLICANT: Finer, Jeffrey  
APPLICANT: Sakowicz, Roman  
APPLICANT: Wood, Kenneth  
TITLE OF INVENTION: No. 6440684el motor proteins and methods for  
FILE REFERENCE: 1016  
CURRENT APPLICATION NUMBER: US/09/592.054  
CURRENT FILING DATE: 2000-07-20  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: fastseq for windows version 4.0  
SEQ ID NO 5  
LENGTH: 1421  
TYPE: DNA  
ORGANISM: Human  
US-09-592-054-5

Query Match 5.6%; Score 33.8; DB 4; Length 1421;  
Best Local Similarity 51.7%; Pred. No. 0.12;  
Matches 77; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 53 TACAGGTGTTCCCAAGAAAGGTGATGACAGTGTGACAGATGAGAGAGTGA 112  
DB 484 TACGAGAGGATCCTTAAGAGGATTAAGATTGTGGACTCAGTGAAGAGACTGTTTGG 543  
QY 113 TACCAACAGCACTGTCAGTGTGGCAATGTGATGCACTATGCAAGCTGAATGAG 172  
DB 544 TTGCTTGTGATAGTGTCTCTGTTTGGACACAGGCGCACACTAGAGTGTGGCTTCA 603  
QY 173 CCACACGGAAGAACATTTCCCTTACCT 201  
DB 604 CGGCTATGAAGTCCAGTGTCCCGATCT 632

Search completed: June 20, 2003, 23:25:48  
Job time : 76.6415 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:58:41 ; Search time 221.698 Seconds  
(without alignments)  
3971.420 Million cell updates/sec

Title: US-09-965-553-11

Perfect score: 600

Sequence: 1 ttggagctggtgcatcata.....tcttagccacaagatctca 600

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	600	100.0	600	10	US-09-965-553-11 Sequence 11, Appl
2	600	100.0	12286	10	US-09-965-553-17 Sequence 17, Appl
3	571.2	95.2	4609	10	US-09-965-553-21 Sequence 21, Appl
4	568.2	94.7	597	10	US-09-965-553-34 Sequence 34, Appl
5	532	88.7	9829	10	US-09-965-553-19 Sequence 19, Appl
6	523.2	87.2	12571	10	US-09-965-553-20 Sequence 20, Appl
7	345.6	57.6	9139	10	US-09-965-553-22 Sequence 22, Appl
8	313.8	52.3	597	10	US-09-965-553-29 Sequence 29, Appl
9	304.6	50.8	600	10	US-09-965-553-27 Sequence 27, Appl
10	172.4	28.7	10482	10	US-09-965-553-23 Sequence 23, Appl
11	80.8	11.5	3813	9	US-10-083-357-634 Sequence 634, App
12	68.4	11.4	631	9	US-10-255-536-16 Sequence 16, Appl
13	64	10.7	4298	9	US-09-997-672-1 Sequence 1, Appl
14	64	10.7	4921	9	US-09-997-672-2 Sequence 2, Appl
15	48.6	8.1	31491	10	US-09-864-680-1 Sequence 1, Appl
16	38.8	6.5	642	9	US-10-123-155-370 Sequence 370, App
17	34.8	5.8	541	9	US-10-184-644-152 Sequence 152, App
18	34.8	5.8	541	9	US-10-184-634-152 Sequence 152, App
19	33.6	5.6	369	10	US-09-960-352-4371 Sequence 4371, App

c	20	33.4	5.6	1355	9	US-10-021-577-1	Sequence 1, Appl1
	21	33	5.5	360	9	US-10-012-896-438	Sequence 438, App
	22	33	5.5	360	9	US-09-895-793-438	Sequence 438, App
	23	33	5.5	360	9	US-09-895-814-438	Sequence 438, App
	24	33	5.5	360	10	US-10-010-940-438	Sequence 438, App
	25	33	5.5	360	10	US-09-759-143-438	Sequence 438, App
	26	33	5.5	360	10	US-09-780-669-438	Sequence 438, App
	27	33	5.5	360	10	US-09-822-827-438	Sequence 438, App
	28	33	5.5	377	10	US-09-960-352-9257	Sequence 9257, App
	29	33	5.5	378	10	US-09-960-352-12818	Sequence 12818, App
	30	33	5.5	379	10	US-09-960-352-9011	Sequence 9011, App
c	31	33	5.5	408	10	US-09-960-352-10425	Sequence 10425, App
	32	33	5.5	425	10	US-09-960-352-2591	Sequence 2591, App
c	33	33	5.5	470	9	US-09-918-995-8924	Sequence 8924, App
	34	33	5.5	2187	9	US-10-175-523-8120	Sequence 110, App
c	35	31.4	5.2	500	10	US-09-925-300-825	Sequence 825, App
	36	31.4	5.2	671	9	US-10-184-644-346	Sequence 346, App
c	37	31.4	5.2	1584	10	US-09-732-224-3	Sequence 3, Appl1
c	38	31	5.2	1629	10	US-09-732-224-6	Sequence 6, Appl1
c	39	31	5.2	809	9	US-09-764-891-2253	Sequence 2253, App
c	40	30.8	5.1	809	9	US-10-205-428-181	Sequence 181, App
c	41	30.8	5.1	472	9	US-09-918-995-33903	Sequence 33903, App
	42	30.6	5.1	699	9	US-10-184-644-138	Sequence 138, App
	43	30.6	5.1	699	9	US-10-184-634-138	Sequence 138, App
	44	30.6	5.1	6271	10	US-09-993-292A-1	Sequence 1, Appl1
c	45	30.6	5.1	6271	10	US-09-993-292A-1	Sequence 1, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-965-553-11  
Sequence 11, Application US/09965553  
Patent No. US2002011259A1  
GENERAL INFORMATION:  
APPLICANT: Wright, David A.  
TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
FILE REFERENCE: P-1065 ISURF Plant Retroelement  
CURRENT APPLICATION NUMBER: US/09/965, 553  
PRIOR FILING DATE: 2001-09-27  
PRIOR APPLICATION NUMBER: 09/322, 478  
PRIOR FILING DATE: 1999-05-28  
PRIOR APPLICATION NUMBER: 60/087125  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 600  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: plant  
US-09-965-553-11  
Query Match 100.0%; Score 600; DB 10; Length 600;  
Best Local Similarity 100.0%; Pred. No. 4,9e-197;  
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TTGGAGGCGGGCTCATATACCCATCTGACAGGCTGGGTAGCCGATGAGGTTG	60
DB	1	TTGGAGGCTGGGCTCATATACCCATCTGACAGGCTGGGTAGCCGATGAGGTTG	60
QY	61	GTTCACAGAAAGGTGATGACAGTGTACGAGATGAGAGATGACTGATACACACA	120
DB	61	GTTCACAGAAAGGTGATGACAGTGTACGAGATGAGAGATGACTGATACACACA	120
QY	121	CGAATGTGCTGCTGTTGGGATGTGTATGCAATGCAAGTGTGATGACGACGCG	180
DB	121	CGAATGTGCTGCTGTTGGGATGTGTATGCAATGCAAGTGTGATGACGACGCG	180



[illegible]

```

RESULT 4
US-09-965-553--34
; Sequence 34, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Glycine max
; US-09-965-553--34

```

Query Match	94.7%;	Score 568.2;	DB 10;	Length 597;
Best Local Similarity	97.0%;	Pred. No. 4.9e-186;		
Matches 579; Conservative	0;	Mismatches 18;	Indels 0;	Gaps 0;

QY	I	TTGGAGGCTGGGCTCATATACCCCATCTCTGACAGCGCTTGGGTAAGCCCAAGTACAGGTG	60
	1	TTGGAGGTTGGGGCTCATATACCCCATCTCTGACAGCGCTTGGGTAAGCCCAAGTACAGGTG	60
Db	1	TTGGAGGTTGGGGCTCATATACCCCATCTCTGACAGCGCTTGGGTAAGCCCAAGTACAGGTG	60
QY	61	GTTCGCCAAGAAAGGTGGAATGACAGCTGTGACAGATAGAGGAAATGACTTGATCCACA	120
	61	GTTCGCCAAGAAAGGTGGAATGACAGCTGTGACAGATAGAGGAAATGACTTGATCCACA	120
Db	61	GTTCGCCAAGAAAGGTGGAATGACAGCTGTGACAGATAGAGGAAATGACTTGATCCACA	120
QY	121	CGAATCTCAGCTGGTGGCAATGTGTATGCATATGCACAAGCTGAATGAAGCCACACGG	180
	121	CGAATCTCAGCTGGTGGCAATGTGTATGCATATGCACAAGCTGAATGAAGCCACACGG	180
Db	121	CGAATCTCAGCTGGTGGCAATGTGTATGCATATGCACAAGCTGAATGAAGCCACACGG	180
QY	181	AAGGACCATTTCCCTTACCTTTTCATGGATCAGATGCTGAGAGACTTGCAGGGCAGGCA	240
	181	AAGGACCATTTCCCTTACCTTTTCATGGATCAGATGCTGAGAGACTTGCAGGGCAGGCA	240
Db	181	AAGGACCATTTCCCTTACCTTTTCATGGATCAGATGCTGAGAGACTTGCAGGGCAGGCA	240
QY	241	TACTACTGTTTCTTGATGATGATCTGGGATPACAAACAGATCGCGGTAGACCCAGAGAT	300
	241	TACTACTGTTTCTTGATGATGATCTGGGATPACAAACAGATCGCGGTAGACCCAGAGAT	300
Db	241	TACTACTGTTTCTTGATGATGATCTGGGATPACAAACAGATCGCGGTAGACCCAGAGAT	300
QY	301	CAGGAGAAGACGGCTTTACATGCGCCCTTTGGCGCTCTTGTCTPACAAGAATGCCATTG	360
	301	CAGGAGAAGACGGCTTTACATGCGCCCTTTGGCGCTCTTGTCTPACAAGAATGCCATTG	360
Db	301	CAGGAGAAGACGGCTTTACATGCGCCCTTTGGCGCTCTTGTCTPACAAGAATGCCATTG	360

OY	361	GGGTTATGTAATGACACACGCCACATTTCAGAGGTGATGCTGCCCATTTTTTCAGACATG	420
Db	361	GGGTATGTATGTAATGACACGCCACATTTCAGAGGTGATGCTGCCCATTTTTTCAGACATG	420
OY	421	GTGGAAGAAAGCATCGAGGTATTTATGACAGCATCTCCGTTTTTGACCCTCATTTTGAC	480
Db	421	GTGGAAGAAAGCATCGAGGTATTTATGACAGCATCTCCGTTTTTGACCCTCATTTTGAC	480
OY	481	AGCTTTTGAGGACACCTAGACAGGGTACTTCAAGAGTGGGAAAGACATACTTGGTACTG	540
Db	481	AGCTTTTGAGGACACCTAGACAGGGTACTTCAAGAGTGGGAAAGACATACTTGGTACTG	540
OY	541	AATGGGAAAAGTGTCAATTCATTCATGTTGTCAGAGGGCATGTGCTTAGGCCAACAAAGATC	597
Db	541	AATGGGAAAAGTGTCAATTCATTCATGTTGTTGTCAGAGGGCATGTGCTTAGGCCAACAAAGATC	597

```

RESULT 5
US-09-965-553-19
? Sequence 19, Application US/09965553
? Patent No. US20020112259A1
? GENERAL INFORMATION:
? APPLICANT: Wright, David A.
? APPLICANT: Voytas, Daniel F.
? TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
? FILE REFERENCE: P-1065 ISORE Plant Retroelement
? CURRENT APPLICATION NUMBER: US/09/965,553
? CURRENT FILING DATE: 2001-09-27
? PRIOR APPLICATION NUMBER: 09/322,478
? PRIOR FILING DATE: 1999-05-28
? PRIOR APPLICATION NUMBER: 60/087125
? PRIOR FILING DATE: 1998-05-29
? NUMBER OF SEQ ID NOS: 41
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 19
? LENGTH: 9829
? TYPE: DNA
? ORGANISM: Glycine max
US-09-965-553-19

```

Query Match	88.7%;	Score.532;	DB 10;	length 9829;
Best Local Similarity	94.0%;	Pred. No. 6.8e-173;		
Matches 564;	Conservative 0;	Mismatches 35;	Indels 1;	Gaps 1;

QY	61	GTTCCTCAAGAAAGGTGGATGACAGTGGTGTACGACATGTGAGACGATGACCTTATATCCAAACA	120
Db	4271	TTGGAGGCTGGGTTTCATATACCCCATCTCTGTATAGCGGTGGGTAAAGTCCAGTACAGGTG	4330
QY	4331	GTTCCTTAAGAAAGGCGGATATACAGTGGTATCGAAATATGAGAGAAATGACTTATATACCAACA	4390
Db	4391	CGAAGTGTCAATGGTGTGGCGAATGTGTATCGACATATCGCAAGTGGAAATGAAGCCACACAG	4450
QY	181	AAGGACCATTTCCCTTTACCTTTACCTTCATGATGATGCTGGAGACATTTGCAGGGCAGGCA	240
Db	4451	AAGGACCATTTCCCTTTACCTTTACCTTCATGATGATGCTGGAAAGGCTTCGAGGGCAGGCA	4510
QY	241	TACTACTGTTCTTTGGATGATCTCGGGATACAAACAGATCGCGGTATGACCCACAGAT	300
Db	4511	TACTACTGCTT-TTGGATGATATTTACGATATACCAACAGATCGCGGTATGACCCACAGAT	4569
QY	301	CAGAGAGAAGCGGCTTTACATGGCCCTTGGGCTCTTGGTTACAGAAAGATCCCATTC	360
Db	4570	CAGAGAGAAGCGGCTTTACATGGCCCTTGGGCTCTTGGTTACAGAAAGATTCATTC	4629
QY	361	GGGTATATGATGACACGACCATTTTCAGAGTGACATGCGGCATTTTTCACAGATG	420
Db	4630	GGGTATATGATGACACGACCATTTTTCAGAGTGACATGCGTACGACATTTTTCAGACATG	4689
QY	421	GTGGAGAAAAGCATCGAGTATTTATGACAGACTTCGTGTTTTTGGACCCCTCATTTGAC	480

Db 4690 GTGAGAGAGACATGAGATATTTATGAGACACTCTGATTTTGGACCCCTCATTTGAC 4749  
OY 481 AGCTGTTAGAGAACTAGAGAGGTACTTGCAGAGTGGAGAGACTAATCTTGAC 540  
Db 4750 AACTATTGAGAACTAGAGATGCTACTAGAGAGTGGTATGACTTGTGACTA 4809  
OY 541 AATTGGAGAAAGTCTATTTATGCTTGCAGAGGCAATGCTAGAGCCCAAGATCTCA 600  
Db 4810 AATTGGAGAAAGTCTATTTATGCTTGCAGAGGCAATGCTAGAGCCCAAGATCTCA 4869

## RESULT 6

US-09-965-553-20  
; Sequence 20, Application US/09965553  
; Patent No. US20020112259A1  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISUR Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/965,553  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 09/322,478  
; PRIOR FILING DATE: 1999-05-28  
; PRIOR APPLICATION NUMBER: 60/087125  
; PRIOR FILING DATE: 1998-05-29  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 12571  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-965-553-20

Query Match 87.2%; Score 523.2; DB 10; Length 12571;  
Best Local Similarity 93.2%; Pred. No. 8.5e-170;  
Matches 559; Conservative 0; Mismatches 38; Indels 3; Gaps 1;

OY 1 TTGGAGGCTGGGCTCATATATACCCATCTCTGACAGCGGTGGTAAGCCAGTACAGTG 60  
Db 4295 TTGGAGGCTGGGCTCATATATACCCATCTCTGACAGCGGTGGTAAGCCAGTACAGTG 4354  
OY 61 GTTCCCAAGAAAGGTGAATGACAGTGTGACAGATGAGAGAAATGACTTGTATACCA 120  
Db 4355 GTTCCCAAGAAAGGTGAATGACAGTGTGACAGATGAGAGAAATGACTTGTATACCA 4414  
OY 121 CGAAGCTGCTACTGTTGGCGCAATGTGTATCGACTATCGCAAGCTGATGAGCCACG 180  
Db 4415 CGAAGCTGCTACTGTTGGCGCAATGTGTATCGACTATCGCAAGCTGATGAGCCACG 4474  
OY 181 AAGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 240  
Db 4475 AAGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 4534  
OY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
Db 4535 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4594  
OY 301 CAGAGAGAGAGCGGCTTTTACATGCGCCCTTTGGCGCTTTGGCTTATGAGAGAGT 360  
Db 4595 CAGAGAGAGAGCGGCTTTTACATGCGCCCTTTGGCGCTTTGGCTTATGAGAGAGT 4654  
OY 361 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
Db 4655 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4714  
OY 421 GTGGAGAGAAAGCATGAGATATTTATGAGAGCACTTCTGCTTTTGGAGCCCTCATTT 480  
Db 4715 GTGGAGAGAAAGCATGAGATATTTATGAGAGCACTTCTGCTTTTGGAGCCCTCATTT 4774  
OY 481 AGCTGTTTGGAGAACTAGAGAGGCTTCTAGAGGTCGAGAGAGACTTACTTGTACTG 540

Db 4775 A---GTTTGAAGAACTAGAGATGCTTATGAGTGTAGAGACTAATCTTGACTG 4831  
OY 541 AATTGGAGAAAGTCTATTTATGCTTGCAGAGGCAATGCTAGAGCCCAAGATCTCA 600  
Db 4832 AACTGGAGAGAGTGTCTACTTGTGCTCAAGAGGCAATGCTTGTAGCCCAAGATCTCA 4891

## RESULT 7

US-09-965-553-22  
; Sequence 22, Application US/09965553  
; Patent No. US20020112259A1  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISUR Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/965,553  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 09/322,478  
; PRIOR FILING DATE: 1999-05-28  
; PRIOR APPLICATION NUMBER: 60/087125  
; PRIOR FILING DATE: 1998-05-29  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 9139  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-965-553-22

Query Match 57.6%; Score 345.6; DB 10; Length 9139;  
Best Local Similarity 73.5%; Pred. No. 2e-108;  
Matches 441; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

OY 1 TTGGAGGCTGGGCTCATATATACCCATCTCTGACAGCGGTGGTAAGCCAGTACAGTG 60  
Db 1646 TAGGAGCAGGCTTATTTACCCCTCTCGATATGATGATGATGATGATGATGATGATG 1705  
OY 61 GTTCCCAAGAAAGGTGAATGACAGTGTGACAGATGAGAGAAATGACTTGTATACCA 120  
Db 1706 GTTCCCAAGAAAGGTGAATGACAGTGTGACAGATGAGAGAAATGACTTGTATACCA 1765  
OY 121 CGAAGCTGCTACTGTTGGCGCAATGTGTATCGACTATCGCAAGCTGATGAGCCACG 180  
Db 1766 AGGACCTGCTACCGGCTGAGAGATGATGATGATGATGATGATGATGATGATGATG 1825  
OY 181 AAGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 240  
Db 1826 AAGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 1885  
OY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
Db 1886 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1945  
OY 301 CAGAGAGAGAGCGGCTTTTACATGCGCCCTTTGGCGCTTTGGCTTATGAGAGAGT 360  
Db 1946 CAGAGAGAGAGCGGCTTTTACATGCGCCCTTTGGCGCTTTGGCTTATGAGAGAGT 2005  
OY 361 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
Db 2006 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2065  
OY 421 GTGGAGAGAAAGCATGAGATATTTATGAGAGCACTTCTGCTTTTGGAGCCCTCATTT 480  
Db 2066 GTGGAGAGAAAGCATGAGATATTTATGAGAGCACTTCTGCTTTTGGAGCCCTCATTT 2125  
OY 481 AGCTGTTTGGAGAACTAGAGAGGCTTCTAGAGGTCGAGAGAGACTTACTTGTACTG 540  
Db 2126 GGGGCTTATTAATCTTGAAGAGTATTAAGAGATGTAAGAGTCAATCTAGTCTC 2185  
OY 541 AATTGGAGAAAGTCTATTTATGCTTGCAGAGGCAATGCTAGAGCCCAAGATCTCA 600  
Db 2186 AATTGGAGAAAGTCTATTTATGCTTGCAGAGGCAATGCTAGAGCCCAAGATCTCA 2245

RESULT 8  
US-09-965-553-29  
; Sequence 29, Application US/09965553  
; Patent No. US20020112259A1  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/965,553  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 09/322,478  
; PRIOR FILING DATE: 1999-05-28  
; PRIOR APPLICATION NUMBER: 60/087125  
; PRIOR FILING DATE: 1998-05-29  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 597  
; TYPE: DNA  
; ORGANISM: Pisum sativum  
US-09-965-553-29

Query Match 52.3%; Score 313.8; DB 10; Length 597;  
Best Local Similarity 70.4%; Pred. No. 5e-98;  
Matches 420; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 1 TTGAGGCTGGGCTCATATACCCCATCTGTACAGCGGCTGGGTAAGCCAGTACAGTG 60  
DB 1 TTGATGGGAGATGATTTACCCGATCTGGATAGTCCATGGGTCCAGTCCCGTCAAGTG 60  
QY 61 GTTCCCAAGAAAGGTGGAAATGACAGTGTACAGATGAGAGGAGTATGATACCAACA 120  
DB 61 GTTCCGAAGAAAGGTGGAAATACCGTATCCGGAAGACAGAGATGATATCCCTACC 120  
QY 121 CGAATGTACATGTTGGGCAATGTGTATGACTATCGAAGCTGATGAAGCAGCAGG 180  
DB 121 AAGTTGCAACGGGGTGGAGATGTGTATGATATAGCGGTTGAATACCGCAACGCA 180  
QY 181 AAGGACATTTCCCTTACCTTTCATGATCAGATCCTGGAGAGACTTGCAGGGCAGCA 240  
DB 181 AAGGACATTTTCCACTCCGTTTCATGATCAGATCAGATCGTGAAGAACTCTCCGGCACA 240  
QY 241 TACTATGTTTCTTGATGGCTATTCGCGGTATACCAATTCGCGTGAACCCGCCGAT 300  
DB 241 TACTATGTTTCTTGATGGCTATTCGCGGTATACCAATTCGCGTGAACCCGCCGAT 300  
QY 301 CAGGAGAGAGCGCTTTACATGCCCCCTTGGCGTCTTTCCTTACAGAAAGATGCCATT 360  
DB 301 CATTAAGAGCGCTTTACATGCCCCCTTGGCGTCTTTCCTTACAGAAAGATGCCATT 360  
QY 361 GGGTATATGATGACACACCATTTACAGAGTGCATGCTGGCCATTTTTCAGAGATG 420  
DB 361 GGGTATGATGACACACCATTTTCACAACGATGTGCAAGCCATTTTTCGCGCACTT 420  
QY 421 GTGGAGAAAGCATGAGGTATTTATGACGACTTTCGCTTTTGGACCCCTCATTTGAC 480  
DB 421 AATGGAAAGCATGAGGTATTTATGACGACTTTCGCTTTTGGTGTATCTTTAGT 480  
QY 481 AGCTTTGAGGAACTTAGAGAGTACTTCAGAGGTGCAGAGGAGTACTTGGTACTG 540  
DB 481 TTATGCTTGGCAAACTTGAAAAGCGTGTGAAAAGATGTGTGAAGACCAATCTTGTGCTT 540  
QY 541 AATTGGAAAAGTGTCAATTCATGTTGAGAGGGGATAGCTTACGCCACAAGATC 597  
DB 541 AATTGGAAAAGTGTCAATTCATGTTGAGAGGGGATAGCTTACGCCACAAGATC 597

Patent No. US20020112259A1  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/965,553  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 09/322,478  
; PRIOR FILING DATE: 1999-05-28  
; PRIOR APPLICATION NUMBER: 60/087125  
; PRIOR FILING DATE: 1998-05-29  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 27  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-965-553-27

Query Match 50.8%; Score 304.6; DB 10; Length 600;  
Best Local Similarity 69.3%; Pred. No. 7.7e-95;  
Matches 415; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 2 TGGAGGCTGGGCTCATATACCCCATCTGTACAGCGGCTGGGTAAGCCAGTACAGTG 61  
DB 2 TTGATGCTGGTGTACATACCCCATCTGTATGATGTTGGTTCCTCCAGTGCATGGCG 61  
QY 62 TTCCCAAGAAAGGTGGAAATGACAGTGTACAGATGAGAGATGACTTGTATACCAAC 121  
DB 62 TTCCCAAGAAAGGTGGAAATGACTGTGTCAAAAATGAAAAGATGACTGCTTACTA 121  
QY 122 GAATGTCTGCTGTTGGCGAATGTGTATGACTATGCAAGCTGATGAAGAACCCACAG 181  
DB 122 GAATGTCTGCTGTTGGCGAATGTGTATGACTATGCAAGCTGATGAAGAACCCACAG 181  
QY 182 AGGACATTTCCCTTACCTTTCATGATCAGATGCTGGAGAGACTTGCAGGGCAGGAT 241  
DB 182 AGGACATTTTCTTTCATCATTTCATGATGACCAAAATGCTTGAACGTTTGGCTATATCAT 241  
QY 242 ACTACTGTTTCTTGATGATGATCTCGGATATCAACAGATCGCGTAGACCCAGAGATC 301  
DB 242 ATTATGCTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301  
QY 302 AGGAGAAAGCGGCTTTTACATGCCCCCTTGGCGTCTTTCCTTACAGAAAGATGCCATT 361  
DB 302 AAGAAAAGCACTTTCACAGTGTCTTATGAACTTTTGCCTATTAAGAAAGATGCCATT 361  
QY 362 GGTATATGATGACACACCATTTACAGAGTGCATGCTGGCCATTTTTCAGACATG 421  
DB 362 GTTATGCAATGCTCTTTCACACATTTTCAGAGTGTATGACTCTATATTTTCAGACTTA 421  
QY 422 TGGAGAAAGCATGAGGTATTTATGACGACTTTCGCTTTTGAACCCCTCATTTGACA 481  
DB 422 TCGAGAGATGAGGAGGTTTTCATGAGCATTTTTCGCTTTCGCTTTCCTTTCCT 481  
QY 482 GCTGTTTGAAGAACTTAGAGAGGTTTATGAGGTGCGAAGAGACTTACTGTACTGA 541  
DB 482 CATGTTTGTGATGTTGCAAGGTTATGCTGTGTTGCGAAGAGACCAATCTTGTCTCA 541  
QY 542 ATTGGAAAAGTGTCAATTCATGTTGAGAGGGGATAGCTTACGCCACAAGATC 600  
DB 542 ATTGGAAAAGTGTCAATTCATGTTGAGAGGGGATAGCTTACGCCACAAGATC 600

RESULT 9  
US-09-965-553-27  
; Sequence 27, Application US/09965553  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto

```

; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 10482
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-23

```

Query Match 28.7%; Score 172.4; DB 10; Length 10482;  
Best Local Similarity 75.2%; Pred. No. 1.8e-48;

Matches 215; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

QY 20 ACCCATCTCTGACAGCGCTTGGGTAGCCCACTAGACAGGTGGTCCCAAGAAAGTGAA 79
DB 4004 ACCCATCTCTGACAGCGCTTGGGTAGCCCACTAGACAGGTGGTCCCAAGAAAGTGAA 4063
QY 80 TGACAGTGGTACGAGATGAGAGGAATGACTTGATACCAACAGAACTGTCTACTGTGGC 139
DB 4064 TGACAGTGGTACGAGATGAGAGGAATGACTTGATACCAACAGAACTGTCTACTGTGGC 4123
QY 140 GAATGTATGACTATGACAGCGCTTGGGTAGCCCACTAGACAGGTGGTCCCAAGAAAGTGAA 199
DB 4124 GAATGTATGACTATGACAGCGCTTGGGTAGCCCACTAGACAGGTGGTCCCAAGAAAGTGAA 4183
QY 200 CTTTCATGATGACAGTCTGAGAGACTTGCAGAGGCGAGCACTACTGTTCTTGATG 259
DB 4184 CTTTCATGATGACAGTCTGAGAGACTTGCAGAGGCGAGCACTACTGTTCTTGATG 4243
QY 260 GATACCTGGGATNACCAAGATCCCGGAGACCCCAAGATCAGCA 305
DB 4244 GATACCTGGGATNACCAAGATCCCGGAGACCCCAAGATCAGCA 4289

```

RESULT 11

US-10-083-357-634

; Sequence 634, Application US/10083357  
; Publication No. US20030054370A1

; GENERAL INFORMATION:

; APPLICANT: Qilandong Zeng et al.

; TITLE OF INVENTION: Systemic Discovery of New Genes

; FILE REFERENCE: 032796-090

; CURRENT APPLICATION NUMBER: US/10/083,357

; CURRENT FILING DATE: 2002-02-27

; NUMBER OF SEQ ID NOS: 1346

; SEQ ID NO 634

; LENGTH: 3813

; TYPE: DNA

; ORGANISM: Saccharomyces cerevisiae

US-10-083-357-634

Query Match 13.5%; Score 80.8; DB 9; Length 3813;  
Best Local Similarity 50.2%; Pred. No. 5.3e-17;

Matches 230; Conservative 0; Mismatches 222; Indels 6; Gaps 1;

```

QY 139 CGAATGTGATGACATTCGCAAGCTGATGAAAGCCACAGGAGAACATTTCCCTTA 198
DB 1108 CGAATGTGATGACATTCGCAAGCTGATGAAAGCCACAGGAGAACATTTCCCTTA 1167
QY 199 CTTTCATGATGATGATGCTGAGAGACTTGCAGGCGAGGAGATCTACTGTTCTTGAT 258
DB 1168 CTTTCATGATGATGATGCTGAGAGACTTGCAGGCGAGGAGATCTACTGTTCTTGAT 1227
QY 259 GGATACCTGGGATACACAGGATCGGGTAGACCCCAAGATCAGAGAAAGAGCGCTTT 318
DB 1228 TTGCAATAGTGTATACACAGGATCCGATGGAACCCAAAGACGCTTCAAAAACGCTTT 1287

```

```

QY 319 ACATCCCCCTTTGGCGTCTTGTCTACAGAGATGCCATTCGGCTATGATGACCA 378
DB 1288 GTCAACCCATCCGGTAACTATGAAATATACCCGTACCATTTGGCTATGATGACCACT 1347
QY 379 GCCAATTTTCAGAGTGCATGCTGGCCATTTTTCAGACATGCTGAGAAAGATCGAG 438
DB 1348 AGTACATTCGCAAGATATACATGCTGATACATTTAGAGACT-----GAGATTCGTCAAT 1401
QY 439 GTATTTATGACGACTTCTCGGTTTGGACCCCTCATTTGACAGCTGTTTGAGAACCTA 498
DB 1402 GTTATCCCTGATGATATATATATATTCCTCCGAATCTCCAGAAAGAACATTTGAA 1461
QY 499 GAGAGGCTACTTTCAGAGTGGAGAGACTTACTGTGACATGATGGAAAGTGTCAAT 558
DB 1462 GACACGCTACTAGAAAGATTAAAGACGAGAACCTCATTTGTAAGAAAGAAATGTAAA 1521
QY 559 TTCATGCTGAGAGGCGCATAGTCTTACGCGCCACAGAT 596
DB 1522 TTTCATCTGTGAAGAACTGAGTTTGTAGGCTATAGTAT 1559

```

RESULT 12

US-10-255-536-16

; Sequence 16, Application US/10255536  
; Publication No. US20030087807A1

; GENERAL INFORMATION:

; APPLICANT: Greenspan, Ralph J.

; TITLE OF INVENTION: Methods for Identifying Compounds for

; FILE REFERENCE: Motion Sickness, Vertigo and Other Disorders Related to

; CURRENT APPLICATION NUMBER: US/10/255,536

; PRIOR FILING DATE: 2002-09-25

; PRIOR APPLICATION NUMBER: US/09/669,751

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: US 60/168,579

; NUMBER OF SEQ ID NOS: 261

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 16

; LENGTH: 631

; TYPE: DNA

; ORGANISM: Drosophila

US-10-255-536-16

Query Match 11.4%; Score 68.4; DB 9; Length 631;  
Best Local Similarity 46.6%; Pred. No. 4.1e-13;

Matches 219; Conservative 0; Mismatches 251; Indels 0; Gaps 0;

```

QY 124 ACTGTCACCTGTTGGCAATGTGTATCGACTATCGCAAGCTGAATGAAGCCACAGGAG 183
DB 105 ACGGAAAAAAGATGGGATAGACAGTTGACTATTCGCAATTAATTAAGAAACTTTTCA 164
QY 184 GACCAATTCCTTACCTTTATGATGATCAGATCAGATCGTGAGAGACTTGCAGGCGCATAC 243
DB 165 GACCAATTCCTTACCTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 224
QY 244 TACGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 303
DB 225 TTTTCATGCTGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 284
QY 304 GAGAGACGCGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
DB 285 GATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 344
QY 364 TTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 423
DB 345 CTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 404
QY 424 GAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483
DB 405 CCATGCAAGCATTTCTATATATGATGATGATGATGATGATGATGATGATGATGATGAT 464

```



[illegible]

OY 525 GACTACTGCTACTGAATTGGAAAAGTGCATT 560  
| | | | | | | | | | | | | | | | | |  
Db 696 AAATAGTCTTTTGGCAATAGGGAATAGTACCTT 731

## RESULT 15

US-09-864-680-1/c  
; Sequence 1, Application US/09864680  
; Patent No. US20020012981A1  
; GENERAL INFORMATION:  
; APPLICANT: Staskiewicz, et al.  
; TITLE OR INVENTION: Bst Resistance Gene  
; FILE REFERENCE: 50687  
; CURRENT APPLICATION NUMBER: US/09/864,680  
; CURRENT FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: 09/360,186  
; PRIOR FILING DATE: 1999-07-23  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 31491  
; TYPE: DNA  
; ORGANISM: Capsicum annuum  
US-09-864-680-1

Query Match 8.1%; Score 48.6; DB 10; Length 31491;  
Best Local Similarity 46.8%; Pred. No. 2.1e-05;

Matches 153; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

OY 139 CGAATGTCATCGACTTCGCAAGCTGATGAAAGCCACACGAGAGACCATTCCCTTA 198  
| | | | | | | | | | | | | | | | | |  
Db 24307 CGGATGTGATAGACTACCGCTGATGAAATGAGATCATGATTAATAATATATCCTCT 24248  
| | | | | | | | | | | | | | | | | |  
OY 199 CCTTCATGGATCGATGCTGGAGAGACTTCAGGGGCGCATACTACTGTTCTTGAT 258  
| | | | | | | | | | | | | | | | | |  
Db 24247 CCTAGGATTCATGACCTTTTGACCAAGCTTCAGGGTCTAAGTCTTTCAAAAATAGAC 24188  
| | | | | | | | | | | | | | | | | |  
OY 259 GGATCTCGGGATACAAACAGATCGCGTAGACCCAGATCAAGAGAGAGAGCGCTTT 318  
| | | | | | | | | | | | | | | | | |  
Db 24187 CTTCGTTCGGGTTCACATGAGTTGAAATTTAGGGAGTCAACATACCAAGACAGCCTTC 24128  
| | | | | | | | | | | | | | | | | |  
OY 319 ACATGCCCCCTTGGCGCTTGTCTACAGAGATGCCATTGGGTTATGTAATGCACCA 378  
| | | | | | | | | | | | | | | | | |  
Db 24127 CGAACCAGATGCTCACTCAAAATTTTATGTCATGTCCTTCGGGTGAGTAACGCCCT 24068  
| | | | | | | | | | | | | | | | | |  
OY 379 GCCAATTTTCAGAGbTGCATGCTGGCCATTTTTCAGACATGTCGAGAGAAAAGCATCGAG 438  
| | | | | | | | | | | | | | | | | |  
Db 24067 GCAGCCTTCATGATCTTATGAGTAGAGTGTCCGTCATTTATGACTTGTTCGTCAAT 24008  
| | | | | | | | | | | | | | | | | |  
OY 439 GATTTATGAGAGCATTCGCTGCTTTT 465  
| | | | | | | | | | | | | | | | | |  
Db 24007 GATTTATGATATATATTCATCTAT 23981  
| | | | | | | | | | | | | | | | | |

Search completed: June 21, 2003, 00:34:41  
Job time : 222.698 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:49:21 ; Search time 2303.77 Seconds  
(without alignments)  
4217.989 Million cell updates/sec

Title: US-09-965-553-11  
Perfect score: 600  
Sequence: 1 ttggaagctggcctcatata.....tcttagccacaagatctca 600

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

- Database :
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estlin:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hlc:\*
  - 9: gb\_estl:\*
  - 10: gb\_est2:\*
  - 11: gb\_hlc:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pln:\*
  - 21: em\_gss\_vit:\*
  - 22: em\_gss\_fun:\*
  - 23: em\_gss\_mam:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_other:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rnd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	317	52.8	582	17	BH421575 BOHEX22TF
2	315.2	52.5	727	17	BH423185 BOHHA57TR
3	315.2	52.5	841	17	BH718174 BOHGI87TR
4	310.4	51.7	759	17	BH655348 BOMEJ58TR
5	310.4	51.7	834	17	BH435343 BOHAU19TR
6	308.8	51.5	762	17	BH497441 BOHBD63TR

C.	7	307.6	51.3	842	17	BH448511 BOGOM52TR
	8	307.2	51.2	768	17	BH714463 BOMPO54TR
	9	307.2	51.2	814	17	BH697068 BOMOT46TR
	10	304.6	50.8	805	17	BH597975 BOHGI95TR
	11	304.4	50.7	854	17	BH711455 BOHTR83TR
	12	302.4	50.4	743	17	BH703456 BOMGU48TR
	13	302.4	50.4	796	17	BH501651 BOGOM29TR
	14	302	50.3	806	17	BH443545 BOGGA22TR
	15	301.6	50.3	697	14	BO996483 OGG12P10.
	16	300.4	50.1	859	17	BO420759 BOG1223TR
	17	297.2	49.5	821	17	BH656439 BOMTR94TR
	18	296.8	49.5	802	17	BH434092 BOGZS83TR
	19	296.6	49.4	576	17	BH530743 BOHOG52TR
	20	294.4	49.0	759	17	BH682723 BOMER67TR
	21	294	49.0	766	17	BH550351 BOHAK33TR
	22	293.4	48.9	806	17	BH244881 AUITA12TR
	23	292.6	48.8	843	17	BH448694 BOGZ15TR
	24	290.8	48.5	799	17	BH701377 BOMHH20TR
	25	290.8	48.5	843	17	BH554612 BOGUM67TR
	26	290.2	48.4	661	17	BH505270 BOHSE73TR
	27	289.8	48.3	854	17	BH452427 BOGNG24TR
	28	288.8	48.1	734	17	BH576493 BOHMB67TR
	29	288.4	48.1	667	17	AQ956471 LERAK19TR
	30	287.8	48.0	757	17	BH654786 BOMDG77TR
	31	287.6	47.9	866	17	BH576412 BOHRE21TR
	32	286.8	47.8	845	17	BH580975 BOGDR46TR
	33	286	47.7	681	17	BH508673 BOGPR31TR
	34	286	47.7	840	17	BH536770 BOHRN81TR
	35	285.2	47.5	744	17	BH405540 Gm_15B001
	36	284.4	47.4	744	17	BH661378 BOMNH43TR
	37	284.4	47.4	811	17	BH588426 BOHAA22TR
	38	282.8	47.1	800	17	BH543949 BOGUR87TR
	39	282	47.0	757	17	BH734399 BOMIH73TR
	40	280.4	46.7	825	17	BH479539 BOGY170TR
	41	279.4	46.6	741	17	BH694557 BOHY120TR
	42	279.4	46.6	760	17	BH484943 BOGL132TR
	43	278.8	46.5	741	17	BH589626 BOGPC86TR
	44	278.4	46.4	737	17	BH709535 BOMAG18TR
	45	278.2	46.4	607	17	BH523876 BOHTW89TR

ALIGNMENTS

RESULT 1  
BH421575/c 582 bp DNA linear GSS 12-DEC-2001  
LOCUS BOHEX22TF BOHE Brassica oleracea genomic clone BOHEX22, DNA  
DEFINITION  
ACCESSION BH421575  
VERSION BH421575.1 GI:17607303  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 582)  
AUTHORS Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.  
TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNML Unpublished (2001)  
COMMENT Other GSSs: BOHEX22TR  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: Sheared ends.  
FEATURES  
source  
1..582  
Location/Qualifiers



KEYWORDS	GSS.
SOURCE	Brassica oleracea.
ORGANISM	Brassica oleracea.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
AUTHORS	Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE	Whole genome shotgun sequencing of Brassica oleracea
JOURNAL	Unpublished (2001)
COMMENT	Other-GSSs: BOMG187TR Contact: Chris Town TIGR
FEATURES	9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: TF Class: sheared ends.
SOURCE	Location/Qualifiers 1..841 /organism="Brassica oleracea" /strain="T01000DH3" /db.xref="taxon:3712" /clone="BOMG187" /clone_id="Bo_2_3_KB" /note="Vector: pHD1; site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHD1 using BstXI linkers"
BASE COUNT	259 a 189 c 150 g 243 t
ORIGIN	
Query Match	52.5%; Score 315.2; DB 17; Length 841;
Best Local Similarity	70.3%; Pred. No. 4.4e-87;
Matches	422; Conservative 0; Mismatches 178; Indels 0; Gaps 0;
Y	1 TTGGAGCGCTGGGCATATACCCATCTGTCAGACCGCTGGGTAGAGCCAGTACAGTG 60
Db	747 TTAAATCTGGTGTATATCTACCCATCTCAGATTCTAAATGGGTATCTCCGTGCATGTT 688
Y	61 GTTCCCAAGAAAGGTGAGATGACAGTGGTGACAGATGAGAGATGACTGATACACA 120
Db	687 GTGCCAAAAAGGTGTGATTAATGATGATTAATAATGACAGATGATTTATACACA 628
Y	121 CGAATCTACGTGTGGCGAATGTGATGACATGACCAAGTCGAATGAAGCACACAG 180
Db	627 AGAACACTTATTTGGGCATAGAGATGTGATTTACCGCAAACTGAACTGTATCTAGA 568
Y	181 AAGACACATTTCCCTTACCTTTATGATGATAGATGCTGGAGACACTTGCAGGCGACA 240
Db	567 AAGGATATTTTCCACATTCATTTATGATGAGATGCTGGAGAACATTCGAATTCACA 508
Y	241 TACTACTGTTTCTTGATGATGATCTCGGATACAAACAGATCGGGTGAACCCAGAGAT 300
Db	507 TTCTATGTTTCTTGATGATGATTCAGAGATCTTCCAGATCCCATCATCATCATATAT 448
Y	301 CAGGAGAAAGCGGCTTTACATGCCCCCTTTGGCGTCTTGTCTTACAGAAGATGCCATT 360
Db	447 CAGAGAAACAAACATTCACATGTCCTTATGTCACCTTGTGATATGCAAGATGCCATT 388
Y	361 GGGTATGATGATGACACAGCCACATTTTCAGAGGTGATCGTGGCCATTTTTCAGACAT 420
Db	387 GGTATGATGATGATGTCAGCACCTTTCAAGGGTCAAGATGTGATCTTTTGATGATG 328
Y	421 GTGAGAAAAGCATGACGAGTATTTATGAGACACTTCTCGGTTTTGGACCCCATTTGAC 480
Db	327 ATTGAGGATGTTGGGAGAGTGTATGATGATATTTTCCGTATGATCTTCCGTTTCT 268
Y	481 AGCTGTTTGAGAACTAGAGAGGATCTTCAGAGGTGCGAAGATCACTTGGTACTG 540
Db	267 GCTTGTGTTGCAAAATTTGTGCAAGGCTCTAGAGAGATGTGAAGACACCAACCTTGTGCT 208
Y	541 AATTGGGAAAAGTCTATTTCATGCTTGCAGAGGGCATATGTCCTTAGGCCACACATCTCA 600

Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80																				

[illegible]

RESULT 5	
BH435343/c	
LOCUS	BH435343
DEFINITION	834 bp DNA
ACCESSION	BOHAU197f BOHA Brassica oleracea genomic clone BOHAU19, DNA sequence.
VERSION	BH435343
KEYWORDS	BH435343.1 GI:17621064
SOURCE	GSS.
ORGANISM	Brassica oleracea.
	Brassica oleracea

REFERENCE	1 (bases 1 to 834)
AUTHORS	Town,C.D., Van Aken,S., Utterback,T. and Frazer,C.M.
TITLE	Whole genome shotgun sequencing of <i>Brassica oleracea</i>
JOURNAL	Unpublished (2001)
Other_GSSs:	BOHAU197R
COMMENT	

9/12 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: [cdtownet@ig.org](mailto:cdtownet@ig.org)  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.

FEATURES	source	Location/Qualifiers
	1..834	/organism="Brassica oleracea" /strain="T01000DH3" /db_xref="taxon:3712" /clone="BOHAU19" /note="Vector: pHOS1; site_1: BstXI; 2-3 kb sheared into pHOS1 using BstXI linkers"
BASE COUNT	219 a 204 c 179 g 232 t	
ORIGIN		

Query Match	51.7%	Score 310.4	DB 17	Length 834
Best Local Similarity	69.6%	Pred. No. 1.4e-85		
Matches 419	Conservative	0	Mismatches 181	Indels 0
				Gaps 0

QY 1 TTGGAGAGGCTGGGCTCATATACCCCTCTGTGACACCGTTGGGTAAAGCCAGTACAGTG 60  
 Db 687 TTGAGTGCAGGGGTATCTACCCAAATTTTCAGACAGTACTTGGGTGAGGCCGGTTCATGTG 622  
 QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGGTACAGATGAGAGGAATGACTTGTATACCAACA 120  
 Db 627 GTTCTTAAGAAAGGTGGCATCACTGTCAATCAAAATAGAAAGATGAGCTGATTTCCATACC 566  
 QY 121 CGAATGTGCACCTGGTGTGGCGAATGTGTATGCATATGCGAAGCTGAAATGAAAGCCACAGG 180  
 Db 557 AGAACAATCACAGGGCATAGAGATGTGCATTGACATACAGGAAGCTAAACTCAGCCACAAGG 508  
 QY 181 AAGGACCAATTTCCCTTACCTTTTATGGATAGAGATGCTGGAGAACTTGCACAGGCGACGCA 240  
 Db 507 AAGGACCACTTCCCACTTCTTTCAATGACCAAGATGCTGGAGAAAGCTACAGCAACACACCC 448

```

Oy 241 TACTACTGTTTCTTGATGGATGATCCGCGGTTAAACAAGATGCGGGTAGACCCCAAGAT 300
Db 447 TACTATGTTTCTTCATGCGTACTCCGGGTTTTTCAGATTACCATTCTTCAGACGAC 388
Oy 301 CAGGAGAAAGCGGCTTTTACATGCCCCCTTTGGCGTTCCTTACAGAAAGATGCCATTC 366
Db 387 CAAGAGAAAGACAACATTACCTACCTGTCCATACGGTACTTTTGCCTACAGGAGAAATGCCCTTC 328
Oy 361 GGGTTATGTATATGCACCAACCCACATTTTTCAGAGGTGCATGCTGGCCATTTTTCAGACATG 420
Db 327 GGATTGTGCAAATGCTCCTCCCACTTCCAGAGATGCATGATGTCATCTTTTACTGATCTT 268
Oy 421 GTGGGAAAAGCATGAGGATTTTATGAGCAGCTTTCGCTTTTGGACCCCTATTGGAC 480
Db 267 ATTTGGAGCAATTATGAGGCTTTTATGAGCAGATTTTCTCAGTCTACCGCTTCTTACTTAAC 208
Oy 481 AGCTTTTGAGGAGACCTAGAGGGGACTTTCAGAGGTGGGAGAGAGCATTACTTGGTACATG 540
Db 207 GACTCCCTTGCCTAACTATGCAAGGGTGGCGGAAGATGAGAGAGAAAGAACTTGGTGCTA 148
Oy 541 AATTGGAAAAGTGTCAATTTCATTTGTTGCGAGAGGGCATATGCTTAGCCACAAGATCTCA 600
Db 147 AATTGGGAAGTGCATTTTCAATGTGTGAAGAAGTGGCATTTGTCTTGGGTACACAGATATCA 88

```

LOCUS	BH497441	762 bp	DNA	linear	GSS 13-DEC-2001
DEFINITION	BOHBD53TR BOHB <i>Brassica oleracea</i> genomic clone BOHBD63, DNA				
ACCESSION	sequence.				
VERSION	BH497441				
KEYWORDS	BH497441.1	GI:17705545			
SOURCE	GSS.				
ORGANISM	<i>Brassica oleracea</i> .				
	<i>Brassica oleracea</i>				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 762)	Town,C.D., Van Aken,S., Utlarback,T. and Fraser,C.M.	Whole genome shotgun sequencing of <i>Brassica oleracea</i>	Unpublished (2001)	Other_GSSs: BOHAB637P
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				
101				
102				
103				
104				
105				
106				
107				
108				
109				
110				
111				
112				
113				
114				

Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: [cdtown@tigr.org](mailto:cdtown@tigr.org)  
DNA is from a doubled haploid provided by Tom Osborn  
Seq primer: TR  
Class: sheared ends.

```

FEATURES
    source
        location/Qualifiers
            1..762
                /organism="Brassica oleracea"
                /strain="T01000DM3"
                /db_xref="taxon:3712"
                /clone="BOHBD03"
                /clone_11b="BOHB"
                /note="Vector: pHOS1, Site_1: BstXI, 2-3 kb sheared
                genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT
    244 a      126 c      162 g      230 t

```

Query Match	51.5%	Score 308.8	DB 17	Length 762
Best Local Similarity	69.7%	Pred. No. 4.1e-85		
Matches 418; Conservative	0	Mismatches 182	Indels 0	Gaps 0

[illegible]

D	b		179	GTCCAAAGGAGGAGGATTTACGTGGTCAAAAATGTAAGATGACTAATACCACA	238
O	y		121	CGAACGTGCTACTGTTGGCGAATGTGATGCATATTCGCAAGCTGAATGAAGCACACGG	180
D	b		239	AGAACAATATACAGACATATAGATGTGCATATGATTATCGAAAACTTAACTCAGACTCTAGA	298
O	y		181	AAGACCATTTCCCCTTACCTTCATGATATGATAGTCCGAGAGACACTTCGACGAGCAGCA	240
D	b		299	AAABATCTTTTTTCCATTGCCATTCATGATATAAATGCTAGAGAGATTTGCAATTCATCC	358
O	y		241	TACTACTGTTTCTTGATGATGATCTCGGATATACACACGATCGGGGTAGACCCAGAGAT	300
D	b		359	TATATATATTTTCTTGATGGGTATTCAGAGGTCTTCCAAATCCCATATCCAAATATAT	418
O	y		301	CAGAGAAAGACGCGCTTTACATGCCCCCTTTGGCGTCTTGGCTACAGAAAGATGCCATT	360
D	b		419	CAMAGAAAAACGACTTTCACATGCCCTTATGTACCTTGTGCTTAGAAGATGACATTT	478
O	y		361	GCGTTATGTATGACACCAAGCCACATTTACAGAGGTGATGCTGGCCATTTTTCACACATG	420
D	b		479	GGGCTATGTAATGTGTCCAGCCATTTCCAAAGATGACAGATGTATTTCTCGATCTT	538
O	y		421	GTGAGAAAAACATCGAGGTATTTATGAGACACTTCCTCGTTTTGGACCCCTCATTTGAC	480
D	b		539	ATAGAGATGTGTTGGAGGTATTTATGATGATCTTCTGTGTACGGATCTTCGTTTCT	598
O	y		481	AGCTGTTTGAGAACCTAGAGAGGCTACTTCAGAGGTGCGAAGACTATMACTTGSTACTG	540
D	b		599	GCTGTTTGTCCATCTTATGAGGTCTCTCAAGAGATGTGAAGACAAACCTTGCTGTG	658
O	y		541	AATTGGAAAAAGTCTCATTTTCATGTTTCGAGAGGGCATGTCTTAGCCACACATCTTCA	600
D	b		659	AATTGGAGAAATGTCACTTCATGTTAGAGAAAGGATTTGCTTGGACACAGAATTTTCA	718

RESULT 7  
BH448511/c

LOCUS	BH448511	842 bp	DNA	linear	GSS 12-DEC-2001
DEFINITION	BOGOM52TF BOGO Brassica oleracea genomic clone BOGOM52, DNA sequence.				
ACCESSION	BH448511				
VERSION	BH448511.1	GI:17634222			
KEYWORDS	GSS.				
SOURCE	Brassica oleracea.				
ORGANISM	Brassica oleracea				
REFERENCE	Eumariophyta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.				
AUTHORS	Tom,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.				
JOURNAL	Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)				
COMMENT	Other GSSs: BOGOM52TR Contact: Chris Town TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: TF Class: sheared ends.				

FEATURES

SOURCE	location/Qualifiers
	1..842
	/organism="Brassica oleracea"
	/strain="T0100DH3"
	/db_xref="taxon:3712"
	/clone="BOGOM52"
	/clone_1fb="BOGO"
	/note="Vector: pHOSt1. site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOSt1 using BstXI linkers"
	227 a 201 c 170 g 244 t

ORIGIN	Query Match	51.3%	Score 307.6	DB 17	Length 842
	Best Local Similarity	69.9%	Pred. No. 1e-84		
	Matches 415	Conservative 0	Mismatches 179	Indels 0	Gaps 0
QY	7	GCCTGGGCTCATATATACCCATCTCTGACAGCCCTTGGGTAAGCCAGTACAGTGGTCTCC	66		
DB	626	GCAGGGGGTGAATCTACCCCAATTTTCAGACAGCACTTGGGGTACCCGGTTCATGTGGTCTTCT	567		
QY	67	AAGAAGGTGAAATGATACAGTGGTACGATGAGAGATGATGATCTTATACCAACAGCAT	126		
DB	566	AAGAAGGTGGCATACACGTCTATCACAATATGAGAGATGAGCTATTTCTTACCGAACA	507		
QY	127	GTCACGTGGTGGCGAATGTATATGACTATATGCAAGTGAATGAAGCCACACGGAAGAC	186		
DB	506	ATCACAGGGCATAGATGTGATGATGACTATACAGAAAGCTAAATCTGACCAAGAAGGAC	447		
QY	187	CATTTCCTTACCTTTTCATGATGATCAGATGCTGGAGATCTTGCAGGGCAGGATCTAC	246		
DB	446	CACATCCCATCTCTTTCATTGACCAAGATGCTGGAAGACATGACCAACCCCTCTATAC	387		
QY	247	TGTTTCCTTGATGATCTCGGATATACCAACAGATCGGGTATGAGACCCAGAGATCAGAG	306		
DB	386	TGTTTCTTCGATGATGATCTCGGATCTTTCATGATACCATTCATTCAGACAGACCAAGA	327		
QY	307	AAGACGGCTTTTACATGCGCCCTTGGCGTCTTGGCTTACAGAAAGATGCCATTCGGGTTA	366		
DB	326	AAGACATTTTACATGCGCTTGGCGTCTTGGCTTACAGAAAGATGCCATTCGGGTTA	267		
QY	367	TGTTATGACACAGCCACATTTTCAGAGTGCATGCTGGCCATTTTTCAGACATGTGGAG	426		
DB	266	TGCATATGCTCGTGAACCTTTCAGAGATGATGATGATCTTTCATGATCTTATTTAG	207		
QY	427	AAAGCATCAGATGATTTATGAGACGATCTTCGCGTTTTCAGACCTCATTTGACAGCTGT	486		
DB	206	GACATTAAGAGAGTTTTATGAGACATTTTCAGCTCAGTCTGCTTTCATTTAGCAGCTGC	147		
QY	487	TTGAGGAACCTTAGAGAGGATCTTCAGAGGTGCGAAGAGACTAACTTGGTACTGATTTG	546		
DB	146	CTTCTATATCTGTGCAAGGCTGCGAAGAGATGAGAGAAAGACTTGTATTAATTTG	87		
QY	547	GAAAGTGTCTTTTCATGATGCTGAGAGGGCATAGTCTTAGGGCCACAAATCTCA	600		
DB	86	GAGAAGTGTCTTTCATGATGAGAGAGTGTGTTTGGTCCACAGATATCA	33		
RESULT 8					
BH714463					
LOCUS	BH714463	768 bp	DNA	linear	GSS 20-FEB-2002
DEFINITION	BOMPO54TR BO_2_3_KB Brassica oleracea genomic clone BOMPO54, DNA				
ACCESSION	BH714463				
VERSION	BH714463.1	GI:18808105			
KEYWORDS	GSS.				
SOURCE	Brassica oleracea.				
ORGANISM	Brassica oleracea				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
TITLE	Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.				
JOURNAL	1 (bases 1 to 768)				
COMMENT	Town,C.D., Van Aken,S., Uteirback,T. and Fraser,C.M.				
	Whole genome shotgun sequencing of Brassica oleracea				
	unpublished (2001)				
	Other_GSS: BOMPO54TF				
	Contact: Chris Town				
	TIGR				
	9712 Medical Center Drive, Rockville, MD 20850, USA.				
	Tel: 301-838-3523				
	Fax: 301-838-0208				
	Email: cdtown@tigr.org				
	DNA is from a doubled haploid provided by Tom Osborn.				
	Seq primer: TR				





sequence.  
ACCESSION BH597975  
VERSION BH597975.1, GI:17850427  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
TITLES Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
JOURNAL 1 (bases 1 to 805)  
COMMENT Town, C.D., Van Aken, S., Uteerback, T. and Fraser, C.M.  
The whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOHGL95TF  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1. 805  
/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHGL95"  
/clone\_lib="BOHG"  
/note="Vector: pHS01; site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pHS01 using BstXI linkers"  
BASE COUNT 210 a 178 c 200 g 217 t

Query Match	Best Local Similarity	50.8%;	Score 304.6;	DB 17;	Length 805;
Matches	415;	Conservative	0;	Mismatches 184;	Indels 0;
					Gaps 0;
QY	2	TGAGAGCTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGGTAAAGCCAGTACAGGTG	61		
Db	10	TTGAGTTCGGGGTATGATCTACCCAAATTTACAGAGACTTGGGTGAGCCCGGTTCATGTG	69		
QY	62	TTCCCAAGAAAGGTGGAATGCAGTGGTGTACAGATGAGAGAAATGATTTGATACCAAC	121		
Db	70	TTCTTAAGAAAGGGTGGGATCACTGTCATCACAAAGGAAAGCGTGAATTCCTACCA	129		
QY	122	GAACGTCTACGTGGTGGGATGTGTATGCACTATGCGAAGCTGTAATGAAGCACACGA	181		
Db	130	GAACAGTACAGGGCATAGATGTGTCATGTGACTACAGAAAGCTAAATCAAGCCACAGGA	189		
QY	182	AGGACCATTTCCCTTACCTTTCAATGATCGATGCTGGAGAGACTTGCAGGGCAGGCAT	241		
Db	190	AAGACCACTTCCCTCTTCTTCAATTCATTCAGACGATGCTGGAAGAAAGTACCAACCTCCCT	249		
QY	242	ACTACTCTTTCTTGATGATGATACACGGGATTCACACCGATGCGGTATACCCACAGATC	301		
Db	250	ACTACTGTTTTCTGATGAGGCTACTACGCGGTTCTTTCAGATACCCATTCATCCAGACGACC	309		
QY	302	AGGAAAGACGGGCTTTTACATGCCCTTTGGCGTCTTTCCTTACAGAAAGATGCCATTCG	361		
Db	310	AAGAGAGACAACTTCACCTGCCCATTCAGGTACTTTTGGCTACAGAGAAATGCCTTTCG	369		
QY	362	GGTATGTAAATGCACACGCCCATTTTCAGAGAGTGCATCTGGCCATTTTTCAGACATGG	421		
Db	370	GATTGTCAATGCTCTGCGCACCTTTCAGAAATCATGATGTGATTCATCTACTGCTTCA	429		
QY	422	TGGAAGAAAGCATCGAGTATTTATGACGACTTCTCGGTTTTTGGACCCCATTTGACA	481		
Db	430	TTGAGGAGATTATGAGAGGTTTTTATGACGATTTCTCAGTCTACGGTCTTCATTTACG	489		
QY	482	GCTGTTTGAGGACCTTAGAGAGGTTACTTCAGAGGTGCGAAAGACCTAATCTTGTACTGA	541		

Db	490	ACTGCGCTTGC	AAC	TGTGCAAGGTCGGGAAGAATGTGAGAGAGAAAGAACCTGGTGTTAA	543
OY	542	ATTGGGAAAAGT	GTCATTTTCAT	TGGTTGCGAGAGGGCCATAGTCCTAGGCCACAAGATCTCA	600
Db	550	ATTGGGAGAACTG	CATTTTCAT	TGGTGTGAAGAAGATACCATTTGTTGGGTCACAGATATCA	608
RESULT 11					
BH711455/c					
DEFINITION	BH711455		854 bp	DNA	linear GSS 20-FEB-2002
LOCUS	BOHX783TF	Bo_2_3_KB	Brassica oleracea	genomic clone BOHX783,	DNA
ACCESION	BH711455				
VERSION	BH711455.1	GI:18801493			
KEYWORDS	GSS.				
SOURCE	Brassica oleracea.				
ORGANISM	Brassica oleracea.				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
TITLE	Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.				
JOURNAL	1 (bases 1 to 854)				
COMMENT	Town,C.D., Van Aken,S., Uteirback,T. and Fraser,C.M.				
	Whole genome shotgun sequencing of Brassica oleracea				
	Unpublished (2001)				
	Other_GSSs: BOHX783TR				
	Contact: Chris Town				
	TIGR				
	9712 Medical Center Drive, Rockville, MD 20850, USA.				
	Tel: 301-838-3523				
	Fax: 301-838-0208				
	Email: cdtown@tigr.org				
	DNA is from a doubled haploid provided by Tom Osborn.				
	Seq primer: TF				
	Class: sheared ends.				
FEATURES	location/Qualifiers				
source	1..854				
	/organism="Brassica oleracea"				
	/strain="TO1000DH3"				
	/db_xref="taxon:3712"				
	/clone="BOHX783"				
	/clone_lib="Bo_2_3_KB"				
	/note="Vector: PHOSI; Site:1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOSI using BstXI linkers"				
BASE COUNT	226 a 207 c 178 g 243 t				
ORIGIN					
Query Match	50.7%; Score 304.4; DB 17; Length 854;				
Best Local Similarity	69.5%; Pred. No. 1e-83;				
Matches 413; Conservative	0; Mismatches 181; Indels 0; Gaps 0;				
OY	7	GCTGGGCTCATATACCCCATCTC	TGACAGCGCTTGGGTAAGCCAGTACAGTGGTGGTCCC	66	
Db	666	GCAGGGGTGATCTACCCCATTTTC	AGACAGCACCTTGGGTGACCCGGTTCATGTGTCTCT	607	
OY	67	AAGAAAGGTGGAATGACAGTGGT	ACGAGATGAGAGGAATGACTTGATACCAACAGAACT	126	
Db	606	ACGAAGGTGGCATCACCGTCATCA	CAAAATGAGAAAGCTGAGTGATTCCTACCAAGAAC	547	
OY	127	GTCACGTGTTGGCAATGTGTATG	ACTATTCGCAAGTGAAGCCACACGGAAGAC	186	
Db	546	GTCACAGGGCATAGAGATGTGAT	TGACTTACAGGAAGCTCAAACTCAGCCACAAGGAAGAC	487	
OY	187	CATTTCGCCCTTACCTTTTCATG	AGATCAGATGGTGAGAGACTTGGACGGGACAGCAPTAC	246	
Db	486	CACCTTCCACTTTTTCATTTGAC	CACAGATGGTGAGAAAGACTTAGCCAACCCCTTACTAC	427	
OY	247	TGTTTCTTGGATGATACCTCGG	GATPACAAACCGATCGGGTAGAAGCCAGAGATCAGAG	306	
Db	426	TGTTTTTCGATGTACTCTGGG	TTCTTTCAGATPACCATTCATTCACAGACGACCAAGAG	367	
OY	307	AAGACGGCCTTTTACATGCCCC	CTTTGGGCTCTTTCCTACAGAAAGATGCCATTTGGGTTA	366	
Db	366	AAGCAACATTTACACTCTGCTC	ATACGATCTTTTCCCTTACAGAGAAATGCCCTTTGGATYG	307	



[illegible]

Query Match	50.3%	Score 302	DB 17	Length 800
Best Local Similarity	70.4%	Pred. No. 5,6e-83		
Matches 404	Conservative	0	Mismatches 170	Indels 0
			Gaps 0	
BASE COUNT	238 a	178 c	146 g	238 t
ORIGIN				
	/clone.lib="BOG"			
	/note="vector: phos1. site.1: BstXI; 2-3 kb sheared genomic DNA inserted into phos1 using BstXI linkers"			
Query Match	2	TGAGGCTGGGCTCATATATACCCATCTCTGACAGCGCTTGGGTAGACCCAGTACAGTGG	61	
Db	575	TAGAGCGCTGTGTATATATCCATCTCTGTATAGTAATTTGGGTAGCCCTGTATCATGTAG	516	
QY	62	TTCCCAAGAAAGTGGATGACAGCGTACAGATGAGAGAAATGACTTGTATTACCAAC	121	
Db	515	TTCTTCAAGAAAGTGGATTAATCTGTAATTAACAAATGAGAAAAAGTACTGATCCCTACTC	456	
QY	122	GAACTGTCACGTGTGGGCAATGTGTATGCATATCGCAAGCTGCAATGAAACCAACAGCA	181	
Db	455	GAACTGTCACGTGTGGGCAATGTGTATGCATATCGCAAGCTGCAATGAAACCAACAGCA	396	
QY	182	AGGACATTTTCCCTTACCTTTATGATGATCAGATGCTGGAGAGACTTGACAGGACGAT	241	
Db	395	AGGATACATTTTCCATTTACCTTTATGATGATCAGATGCTGGAGAGACTTGACAGGACGAT	336	
QY	242	ACTACTGTTTCTTGATGATGATACACGATGATGATGATGATGATGATGATGATGATGATG	301	
Db	335	ACTATGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	276	
QY	302	AGGAGAGAGCGCCCTTTCATATGCCCCCTTTGGCGCTTTTCTTACAGAAAGATGCCATTGC	361	
Db	275	AGGAGAGAGCGCGCTTTCATATGCCCCCTTTGGCGCTTTTCTTACAGAAAGATGCCATTGC	216	
QY	362	GTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	421	
Db	215	GCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	156	
QY	422	TGAGAAAGAGATCGAGGTATTTATGACGACTTCTCGTTTGTGGAGAGAGATGATGACA	481	
Db	155	TTGAGAGATATATGAGAGGTATTTATGACGACTTCTCGTTTGTGGAGAGAGATGATGACA	96	
QY	482	GCTGTTTGAAG	541	
Db	95	TTTGTGTTTGAAG	36	
QY	542	ATTGGGAAAGTGTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG	575	
Db	35	ACTGGGAGAGAGTGTATTTATGATGATGATGATGATGATGATGATGATGATGATGATG	2	
RESULT 15				
LOCUS	BO996483	697 bp	mRNA	linear
DEFINITION	OG612P10.yg.ab1 OG_EFGHJ lettuce serriola Lactuca sativa cDNA clone			
ACCESSION	BO996483			
KEYWORDS	BO996483.1 GI:22430879			
ORGANISM	Lactuca sativa.			
SOURCE	Lactuca sativa.			
REFERENCE	1 (bases 1 to 697)			
AUTHORS	Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,D., Ellison,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.			
TITLE	Lettuce and Sunflower ESTs from the Compositae Genome Project			
JOURNAL	http://compgenomics.ucdavis.edu/unpublished (2002)			

## COMMENT

Contact: Alexander Kozik (R.W.Michelmores)  
Department of Vegetable Crops, R.W.Michelmores Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-752-9659  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]  
singleton, see http://cgpdb.ucdavis.edu/ for details.  
Plate: OG612 row: P column: 10.  
Location/Qualifiers

## FEATURES

source

1. 697  
/organism="Lactuca sativa"  
/cultivar="L.seriola  
/db\_xref="taxon:4236"  
/clone="OG612P10"  
/clone\_1lb="OG-FRCHJ lettuce seriola"  
/lab\_host="E.coli"  
/note="Vector: pBRCDNA5flab. The library was constructed  
from 10 different sources of RNA from a single genotype.  
Separate cDNAs were generated using primers that  
incorporated unique 5' and 3' tags to distinguish each  
source of RNA. cDNAs were then pooled, size-fractionated,  
directionally cloned into a custom medium-copy vector and  
size bias. Details of each source of RNA and library  
construction can be obtained at http://cgpdb.ucdavis.edu/  
TAG\_TISSUE-flowers environmental stress  
TAG\_SEO-CGATGCCGGG"

BASE COUNT

203 a 112 c 162 g 220 t

ORIGIN

Query Match 50.3%; Score 301.6; DB 14; Length 697;

Best Local Similarity 69.6%; Pred. No. 7e-83;

Matches 409; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
12 GCTCATATACCCATCTCTACAGCCGCTGGGTAAAGCCATACAGGTGTTCCAGAA 71
1 GATTATCTACCCGATTCGATAGTAGTATGATGATGATGATGATGATGATGATGAT 60
72 AGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 131
61 AGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
132 TGGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191
121 AGGCTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
192 CCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251
181 CCCATTACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
252 CTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 311
241 TTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
312 GGCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
301 TACTTTCACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
372 TGCACAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 431
361 TGCACAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
432 CATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 491
421 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
492 GAACCTAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 551
481 AAACCTAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
552 GTGTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 599
```

Db 541 ATGCCATTATGATAAAGAGGATATGCTTGGACACAGGTTTC 588

Search completed: June 20, 2003, 22:39:20  
Job time: 2307.77 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 19:31:16 ; Search time 75.6226 Seconds

(without alignments)  
6927.165 Million cell updates/sec

Title: US-09-965-553-2

Perfect score: 18

Sequence: 1 tggcgccgttgcgggga 18

Scoring table: IDENTITY-NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_da:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_com:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_mus:\*  
34: em\_hlg\_pla:\*  
35: em\_hlg\_rtd:\*  
36: em\_hlg\_mam:\*  
37: em\_hlg\_vrt:\*  
38: em\_sy:\*  
39: em\_hlgo\_hum:\*  
40: em\_hlgo\_mus:\*  
41: em\_hlgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	10128	8	AF186182 Glycine m
2	18	100.0	10510	8	AF186186 Glycine m
3	18	100.0	83339	8	AP001311 Arabidops
4	18	100.0	106716	8	AC006413 Arabidops
5	18	100.0	122116	8	AP004483 Lotus jap
6	18	100.0	148246	8	AP003054 Oryza sat
7	18	100.0	189744	2	AP003577 Oryza sat
8	18	100.0	194640	2	AP005160 Oryza sat
9	18	100.0	261265	8	AF427791 Hordeum v
10	17	94.4	67517	8	AF028277 Hordeum c
11	16.4	91.1	537	8	AX059280 Sequence
12	16.4	91.1	687	6	AX059280 Sequence
13	16.4	91.1	1732	8	BNDNTRNA
14	16.4	91.1	4717	8	AF439379 Medicago
15	16.4	91.1	6676	8	SCRI731
16	16.4	91.1	8117	8	TAGLUTAG
17	16.4	91.1	11094	1	AE005854 Caulobact
18	16.4	91.1	12314	1	PSJ000640 Pisum sat
19	16.4	91.1	15085	1	AE007177 Mycobacte
20	16.4	91.1	15417	1	AE006966 Mycobacte
21	16.4	91.1	18789	8	AF439380 Medicago
22	16.4	91.1	33254	8	AP004537 Lotus jap
23	16.4	91.1	36526	1	MSGY42
24	16.4	91.1	36804	1	MTCCY210
25	16.4	91.1	48128	6	AX059513 Sequence
26	16.4	91.1	62916	8	AB046427 Arabidops
27	16.4	91.1	64165	2	AC067965 Arabidops
28	16.4	91.1	65979	8	AF254799 Hordeum v
29	16.4	91.1	77132	2	AP003820 Oryza sat
30	16.4	91.1	77136	8	AF474982 Hordeum v
31	16.4	91.1	78805	8	OSJN00123
32	16.4	91.1	80930	8	AP004894 Lotus jap
33	16.4	91.1	82532	8	AC006219 Arabidops
34	16.4	91.1	83991	8	AP004965 Lotus jap
35	16.4	91.1	85748	8	OSJN00251
36	16.4	91.1	87219	8	AP002054 Arabidops
37	16.4	91.1	90101	2	AP003807 Oryza sat
38	16.4	91.1	90627	8	AF000411 Arabidops
39	16.4	91.1	90840	8	AB073158 Arabidops
40	16.4	91.1	92281	8	AP004896 Lotus jap
41	16.4	91.1	94139	8	AP003805 Oryza sat
42	16.4	91.1	95681	8	AP002035 Arabidops
43	16.4	91.1	95959	8	AC006217 Arabidops
44	16.4	91.1	98303	2	AP004642 Oryza sat
45	16.4	91.1	100595	8	AP004545 Lotus jap

## ALIGNMENTS

RESULT 1  
LOCUS AF186182 10128 bp DNA linear PLN 05-JAN-2000  
DEFINITION Glycine max retrovirus-like element Calypso-1, partial sequence.  
ACCESSION AF186182  
VERSION AF186182.1 GI:6671122  
KEYWORDS

## SOURCE

Glycine max.  
Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae;

## REFERENCE

1 (bases 1 to 10128)  
Wright,D.A. and Voytas,D.F.

Pred. No. is the number of results predicted by chance to have a

TITLE Calypso: A Heterogeneous Retrovirus-like Element Family from  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 10128)  
 AUTHORS Wright, D.A. and Voytas, D.F.  
 JOURNAL Direct Submission  
 Submitted (14-SEP-1999) Zoology and Genetics, Iowa State  
 University, 2208 Molecular Biology Bldg., Ames, IA 50011, USA  
 FEATURES  
 source  
 1. 10128  
 /organism="Glycine max"  
 /strain="L85"  
 /db\_xref="taxon:3847"  
 repeat\_region  
 294..>10128  
 /rpt\_family="retrovirus-like element Calypso1-1"  
 LTR  
 primer\_bind  
 1605..1623  
 1759..1717  
 /gene="pol"  
 /pseudo  
 8068..8976  
 /gene="envelope-1like"  
 /pseudo  
 BASE COUNT 3031 a 2135 c 2330 g 2632 t  
 ORIGIN  
 Query Match 100.0%; Score 18; DB 8; Length 10128;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 TGGCGCCGTTGTGCGGGA 18  
 ||||||||||||||||  
 1605 TGGCGCCGTTGTGCGGGA 1622  
 ||||||||||||||||  
 Db 1605 TGGCGCCGTTGTGCGGGA 1622  
 ||||||||||||||||  
 RESULT 2  
 AF186186 10510 bp DNA linear PLN 10-SEP-2001  
 LOCUS Glycine max retrovirus-like element Calypso-1, partial sequence.  
 ACCESSION AF186186  
 VERSION AF186186.1 GI:6671126  
 KEYWORDS  
 SOURCE  
 Glycine max.  
 Glycine max.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.  
 1 (bases 1 to 10510)  
 Wright, D.A. and Voytas, D.F.  
 Calypso: A Heterogeneous Retrovirus-like Element Family from Glycine  
 max  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 10510)  
 AUTHORS Wright, D.A. and Voytas, D.F.  
 JOURNAL Direct Submission  
 Submitted (14-SEP-1999) Zoology and Genetics, Iowa State  
 University, 2208 Molecular Biology Bldg., Ames, IA 50011, USA  
 FEATURES  
 source  
 1. 10510  
 /organism="Glycine max"  
 /strain="L85"  
 /db\_xref="taxon:3847"  
 31..>10510  
 /rpt\_family="retrovirus-like element Calypso-1"  
 LTR  
 primer\_bind  
 1478..1495  
 1644..8392  
 /gene="pol"  
 /pseudo  
 10144..10510

/gene="envelope-1like"  
 /pseudo  
 BASE COUNT 3186 a 1844 c 2306 g 3174 t  
 ORIGIN  
 Query Match 100.0%; Score 18; DB 8; Length 10510;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 TGGCGCCGTTGTGCGGGA 18  
 ||||||||||||||||  
 Db 1478 TGGCGCCGTTGTGCGGGA 1495  
 ||||||||||||||||  
 RESULT 3  
 AP001311/c 83339 bp DNA linear PLN 27-DEC-2000  
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MW11.  
 DEFINITION AP001311 BA000014  
 ACCESSION AP001311.1 GI:7209747  
 VERSION  
 KEYWORDS  
 SOURCE  
 Arabidopsis thaliana (strain:Columbia) DNA, clone\_lib:Mitsui P1  
 clone:MW11.  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (sites)  
 Kaneko, T., Katoh, T., Sato, S., Nakamura, A., Asamizu, E. and Tabata, S.  
 Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC  
 and BAC clones  
 DNA Res. 7 (3), 217-221 (2000)  
 20363099  
 2 (bases 1 to 83339)  
 Kaneko, T., Katoh, T., Sato, S., Nakamura, Y., Asamizu, E. and Tabata, S.  
 Direct Submission  
 Submitted (01-MAR-2000) Yasukazu Nakamura, Kazusa DNA Research  
 Institute, Department of Plant Gene Research; 1532-3, Yana,  
 Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,  
 Tel:81-438-52-3935, Fax:81-438-52-3934)  
 Address for correspondence: kaos@kazusa.or.jp  
 For the latest information on annotation of this clone, please see  
 http://www.kazusa.or.jp/kaos/cgi-bin/agd-graph.cgi?c=MW11  
 Genes with similarity to proteins in the databases are described in  
 'product' or 'note' qualifiers. Genes that have no significant  
 protein similarity are described as 'unknown protein'.  
 The software programs used to predict genes include: Grail  
 (Informatics Group, Oak Ridge National Laboratory,  
 http://compbio.ornl.gov/Grail-1.3/),  
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),  
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of  
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and  
 SplicePredictor (Volker Brendel, Stanford University,  
 http://greenlab.zool.iastate.edu/cgi-bin/sp.cgi).  
 Genes encoding tRNAs are predicted by tRNAscan-SE  
 (Sean Eddy, Washington University School of Medicine, St. Louis,  
 http://genome.wustl.edu/eddy/tRNAscan-SE/).  
 This sequence may not be the entire insert of this clone. It may be  
 shorter because we remove overlaps between neighboring submissions.  
 The 5' clone is T6122 and the 3' clone is MSJ3.  
 Location/Qualifiers  
 1. 83339  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="3"  
 /clone="MW11"  
 /clone\_lib="Mitsui P1"  
 complement(join(1709..1889,2017..2172,2316..2444,  
 2612..2790))  
 /note="gene\_id:MW11.1  
 unknown protein"

CDS

```

/|codon_start=1
/|evidence=not_experimental
/|protein_id="BAB02224.1"
/|db_xref="GI:9294327"
/|translation="MLETKITFWEVCQLGFGPIDLPGSOKCRPDYPLALVKKHPN
SGQTVWMSFGKSTLKYSGMDRVWRIRVRYEPKLTLYONIFKLENTQLKEF
EYVRLDVQVQSDLGSGFWLKFRIKRSIDILKFWVRIGLFGSGSVQFEGFEMVYV
DPLVWLESETMEERLVCLRDSMMYKERSKSVKELGVRAEDVLEMSKNPS"
17169..17450
/|note="contains similarity to replication protein A1
gene_id:MVA11.2"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|join(17769..18220,19125..19210,19322..20811,20886..21233,
21298..21477,21554..21653,21734..22124,22201..22517)
/|note="contains similarity to unknown protein
gb|AA06087.1
gene_id:MVA11.3"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|join(23436..23872,24663..24795,24882..25169,25274..25527,
25580..25913)
/|note="gene_id:MVA11.4"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|product="En/Spm-like transposon protein-like"
/|complement(28752..30974)
/|note="gene_id:MVA11.5"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|product="retroelement pol polypotein"
38275..38444
/|note="gene_id:MVA11.6
unknown protein"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|join(40447..41304,41458..42397,42583..43109)
/|note="contains similarity to reverse transcriptase-
gene_id:MVA11.7"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
43611..43908
/|note="gene_id:MVA11.8
unknown protein"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|complement(join(44798..44894,44968..45926,46014..46233))
/|note="gb|AA018631.1
gene_id:MVA11.9
similar to unknown protein"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|complement(join(47335..48223,48359..49241,49413..50566,
50705..51104,51174..51337,51485..52374))
/|note="gene_id:MVA11.10"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|product="gag-protease polypotein-like"
/|complement(join(53267..55657,56051..57537))
/|note="contains similarity to Pol-like peptide
gene_id:MVA11.11"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
complement(join(57568..57707,57779..57928,58051..58165,
58276..58305))
/|note="contains similarity to disease resistance RPP5 like
protein
gene_id:MVA11.12"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|join(59311..59421,59522..59671)
/|note="contains similarity to uroporphyrinogen III
methyltransferase
gene_id:MVA11.13"
/|codon_start=1
/|evidence=not_experimental
/|protein_id="BAB02225.1"
/|db_xref="GI:9294328"
/|translation="MGFKTLPSPKIMDHGLPEFDPAAVERGTPPLNLYVRELKD
FATEIOSAGVSPSPILIGKVELSLPMHCKMSSCLVETQ"
complement(63246..66054)
/|note="gene_id:MVA11.14"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|product="retroelement pol polypotein"
/|complement(join(70141..70333,70737..70900,71054..71187,
71233..71533,71588..72066,72199..72289,72344..72573))
/|note="gene_id:MVA11.15"
/|pseudo
/|codon_start=1
/|evidence=not_experimental
/|product="Athila retroelement protein-like"
/|join(75749..75952,76110..76202,76285..76558,76629..76719,
76801..76930)
/|note="gb|AAD25622.1
gene_id:MVA11.16
strong similarity to unknown protein"
/|codon_start=1
/|evidence=not_experimental
/|protein_id="BAB02226.1"
/|db_xref="GI:9294329"
/|translation="MCIGPFSQKTYEPKPKKTQKLTQIDFIFNSIDAIGRTKE
KKSVAIAPWSPSPFSFSGPSTVGRILREFDANIKKQGFVILLDEKCSF
IHAFIPALASHEROVLRREGIIFENVGFEVGRCTKXIKTIDHDFLPATITIEFS
DVGPRIEREKPKLRPNDFNLOALNTIEIPDVYQGTFFVQSNLNPSTQRLVLR
IDSSVIVTISLDVYATRAHLSGDTTYSVMLITTVNPKMFG"
/|join(77345..77871,78106..79438)
/|note="contains similarity to unknown protein
gb|AAD25621.1
gene_id:MVA11.17"
/|codon_start=1
/|evidence=not_experimental
/|protein_id="BAB02227.1"
/|db_xref="GI:9294330"
/|translation="MEVLDSSTKMYTRRKORENIGRTIVNLTPTAGDLVYLRILNKVK
GATSPDIKTYGVGVHSEPKACDHAGGLDQKEMDAMDDEAOMGSSYLRSLEFVLI
LITCVSEPKLMSHCWESMADVLKQQRVNLFPQLELKAKELEKTLLEIETLQO
HEKSLSDYPEMPQEPKSNKGNVPAVSAIAALLLPGGRASHWFKIPINVHEDFICD
IKIGSLANVLSKVDLIITWDEAPMAHRTFEAVDRILRILISGDEKALRTKLGKIV
LLGCDFOILPVIPTORTQETVSAAINRSYLSQNMARQVPEIKFEAWLIL
OIGDEAPRKTHGIDDOEDENITDKNLILPTEMLLEVLCOSVSDPFNFTODLEL
LKGTAVLPARNETVNEINDYLSKVGCLAKVEFSASIDODEALTEGPFMSVPMETL
NSLEPGLPAHRLCLKGVPTMLRLNLRNKEGICNSTRLTLYTHLGDVKYAKAELISDTT
KRRKRVLPRIILISPODSKHPFTLRRROPVPRVCYMTVNRKSGOTLRVALVLPKRV
FSHGOLYVALSRVTSPPKGLTVLDTSSKKKGYVTNIVREVNGPLAITGKKS"
BASE COUNT 26491 a 15765 c 14391 g 26692 t
ORIGIN
Query Match 100.0%; Score 18; DB 8; Length 83339;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TGGCGCGTGTGCGGGGA 18
|||||

```

Db 31274 TGCGCCGCTTGCGGGA 31257

RESULT 4  
AC006413/c

LOCUS AC006413 106716 bp DNA linear PLN 11-MAR-2002

DEFINITION Arabidopsis thaliana chromosome 2 clone F5K7 map ve013, complete sequence.

ACCESSION AC006413

VERSION AC006413.4 GI:20197765

KEYWORDS HTG

SOURCE Arabidopsis thaliana.

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 106716)  
Lin, X., Kaul, S., Shee, T.P., Fujii, C.Y., Shen, M., VanAken, S.E., Barnstead, M.E., Mason, T.M., Bowman, C.L., Ronning, C.M., Benito, M.-I., Carter, A.J., Cressy, T.H., Buell, C.R., Town, C.D., Nierman, W.C., Fraser, C.M. and Venter, J.C.  
Unpublished

REFERENCE 2 (bases 1 to 106716)  
Lin, X.  
Direct Submission

JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

REFERENCE 3 (bases 1 to 106716)  
Town, C.D. and Kaul, S.  
Direct Submission

AUTHORS

TITLE Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org

COMMENT On Apr 18, 2002 this sequence version replaced gi:6598560.

FEATURES

source location/Qualifiers

1..106716  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="2"  
/map="ve013"  
/clone="F5K7"  
complement(1..5067)  
/note="overlap with BAC clone F18P14 (AC006918:1..5067)."  
891..1187  
/gene="At2g06230"  
/note="F5K7.2: predicted by genSCAN"  
891..1187  
/gene="At2g06230"  
891..1187  
/gene="At2g06230"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AA015238.1"  
/db\_xref="GI:20197765"  
/translation="MQRLEGVTPYSLSCRTWRERMRMEAPKDWVCYCGPKVADIP LPLKMCVETDSSPIFKPWPESDSTGDSGADDFLWRVDNPKNPDTLFO"  
complement(1930..4317)  
/gene="At2g06220"  
/note="F5K7.1: predicted by genSCAN"  
complement(1930..2249,2397..2559,2636..2794,3087..3184,3509..3698,3901..3995,4056..4317))  
/gene="At2g06220"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AA015236.1"  
/db\_xref="GI:20197767"  
/translation="MARVDTGTGTHYESTRPESEVSTQTSSEIFIDGAAVEVP TEPAETDSANKTPKTAETAEIIEESLDGEASKGADIEKELTANEGGDLLE EEMVNTTSGEVNINCODTCEADDDSDGPEDEYTKENEDEDAENSLKEDDDSDS OPLAPETMTFSFEYDKVCKISKIGIVSKVLTYEKIRSGTDRKLVLFLVKILA

gene  
/gene="At2g06210"  
/note="F5K7.3: TPR repeat protein"  
join(<5036..5176,5353..5464,5508..5702,5873..5973,6170..6223,6446..6653,6760..6974,7189..7276,7561..7661,7760..7852,8195..8263,8371..8463,8538..8798,9307..9387,9479..9637,10062..10157,10379..10480,10592..10666,10766..10990,11296..11562,11666..11737,11926..12258,12343..>12516)  
/gene="At2g06210"  
join(5036..5176,5353..5464,5508..5702,5873..5973,6170..6223,6446..6653,6760..6974,7189..7276,7561..7661,7760..7852,8195..8263,8371..8463,8538..8798,9307..9387,9479..9637,10062..10157,10379..10480,10592..10666,10766..10990,11296..11562,11666..11737,11926..12258,12343..12516)  
/gene="At2g06210"  
/codon\_start=1  
/product="putative TPR repeat nuclear phosphoprotein"  
/protein\_id="AA015237.1"  
/db\_xref="GI:20197768"  
/translation="MASYIIPVONSEEEVRYVLDLPDASDILDILKAQAPLML IIRERYGKRIEOPFOILEEGSSGKLYDASICRIDDEYAVKYRIILN ALGAYVYLGKTEKNEKEOFISATRYKASRIDHEPTWGGFLAKGID NALQAEKIVLDTAPDNPALIGQASVEENRPFSSRLQYKGLDKRQADRVIA SCGRMTSSSDINDACRQOYVITITITIFLOLDPNVALVAGINDQANDSICMR GMDMOQAEIETPYCASHLNTLANHFETGTHFLVLEQLERLATVTHGIRKSHENK LANSYHSRGEFEKAGMYMAIKETNNPHFVFPFELGQVQLTGLKGSVENEK VLEAYPNCETLKAIGLTYTOLGQNEKALEEMRKATLDPDPAQAFVIGLILLISDT GAALDAFKMARTLAKKGQEVPIEVLNIGALHFEREFESALENEKEALGIMISF LDEKENLEOGVSVLYKDTGIFRHLISGKSNQNNPLAIEVNEALYKDXDNPNALISLGE ATEFWYRIIEPKYPGYDAVYLRLASAKQNNPLAIEVNEALYKDXDNPNALISLGE LEKNDWYKAKETFRANDTIDGDSITALLSLGWNVFAAMRNKRPNKLEATHLEK AKELTYVLTQHSNMTAAANGSGLVLAQGFNDIKLVFTQVQASAEQWQECRKLTLR VNLAHYFAQGNFALTVMYQNCIRKFEYFNDSQILLYLARTHYAEQWQECRKLTLR AIHLTPSNYTERFDLGAVMOKSSSTLOKRRKADDEVSTVAEAENAVVFTQLSAS DLHYHGFDSKRIOTHOVQCSHLENAKYHRAEQEELONRQLERARQALAEAR KAEBORVLEKRRQEEELRLKQEEEFORLIKDKMSSTQCSKRRKQVREDDCESK PSERRRKKGGRRKKDKSSRRARHYEDDEEATYADNDHNEVEDDEANTNRNDEMTQ EAEPRVDDADHDLAAGLEDDPDVDDDEVPSVGRRRALSSDEGELMEBSHPNWS POKEEKSENGEAGDPNMEEEENAN"  
complement(12724..12818)  
/note="DNA sequence related to Tag1 transposon (GB: L12220)."  
complement(13650..13673)  
/rpt\_family="(GA)n"  
complement(14301..15353)  
/gene="At2g06200"  
/note="F5K7.4: predicted by genSCAN and grail; supported by cDNA: gi:16974616.gb\_Ay060566.1"  
complement(join(<14301..14687,14776..14995,15089..>15353))  
/gene="At2g06200"  
complement(join(14369..14687,14776..14995,15089..15284))  
/gene="At2g06200"  
/codon\_start=1  
/product="expressed protein"  
/protein\_id="AAD19765.2"  
/db\_xref="GI:20197770"  
/translation="MATRIPTESQWELNQALVFKYLAANMPVPHLLFLIKRPL FSSSSSSSSSSSPSLSPHFGNNVYMGMGKRIADAPGRCRDGRKWRCSKAYP DSKYCEHRHMRGKRSSSRKPPPOFTNPLDSSRRRSRGYMDPFSIERSGISKS CSGSAMENDNDGSCRGINNEKOPDRHCFILGTDRTERPLMLEKLIKQRHNDNEE OGSKRTFRFLDEWPSKSSVSTSLFI"  
15121..15159  
/rpt\_family="(GAA)n"  
complement(16633..16694)  
/rpt\_family="AT-rich"  
17318..26761  
/note="Putative Ty3-type retrotransposon, structure: long



terminal repeat A (LTRA), putative Ty3-type pol  
polyprotein-1, putative Ty3-type pol polyprotein-2, long  
terminal repeat B (LTRB)"  
complement (17533..18380)  
/gene="At2g06190"  
/note="F5K7.5"  
complement (join(<17533..17932,17938..>18380))  
/gene="At2g06190"  
complement (join(17533..17932,17938..18380))  
/gene="At2g06190"  
/codon\_start=1  
/product="putative Ty3-gypsy-like retroelement pol  
polyprotein"  
/protein\_id="A019756.1"  
/db\_xref="GI:4388718"  
/translation="MDEVLGIPRQGVDSVFVYVDRSKMTHFACKRTADASNAK  
LFEVYRLHGVKSTSDRDTKFLSHWSTLMRFGALNLSSTPHQIDQTEYTN  
RTLGNNVRSICGDNPKQMDLAPQIEFAYNSVHVDLVLKPKALGASAEYAEILV  
VKEVYKAKLEATGKKNVVAADKRRRFVKEGDDVMVLLRGKFAVGYNKKPKRYG  
PFYLRKINDNAVVALPKSMNISNTENVADLHEHYADGVLYPEENLRTSSEVEED  
VGEI"  
19174..22307  
/gene="At2g06180"  
/note="F5K7.6"  
join(<19174..19325,19483..19702,19717..20174,20300..20483,  
20634..20843,20930..21110,21129..21302,21427..21586,  
21665..>22307)  
/gene="At2g06180"  
join(19174..19325,19483..19702,19717..20174,20300..20483,  
20634..20843,20930..21110,21129..21302,21427..21586,  
21665..22307)  
/gene="At2g06180"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="A019757.1"  
/db\_xref="GI:4388719"  
/translation="MTRLSAPHLTAQIRREESTISKIYVMASLNNPPLHLSIR  
QLTRPDLDDSONKSNHSGDDSSSKMRGASASROPRVITTGGMGPIRSPRV  
DKRKKAPKQKGLSPMTNNDLTVLDSGGGVDPKVAVAQERAGGVLCIROPRE  
MAVLPDSGVHPRALLPRRLSSNVSRFRPRACVITPGLRGEORIRIPRAALISG  
QDQDERPDLVHSCSMISDPNWSGLSGKLLASQILLQGHHSVGDDEPARITRT  
CKDQRMFFDEDTLSVDPSSKLSNRLLAFNDRNTALPIDPADVIALPAQPPLE  
EVPOLYERPSAFQKNEVTSKSCSTSCQTSARAGDSAVVAARTSLAASHTPESV  
SHPSLPANAPLILRRARLLPVYRLAEILVIDEVAISREERKPLANDEAAVS  
DRSSGSGRGRGSRGPRNSQEVROTSNLSAANRPHMSYTHRDHPIEDK  
AGLANLRIHKGRCQVPGDCCTKQVCSKATSKSGFGEARREDPHLCVDSRE  
VRSGARGLPGMLKLMELSTSKDEKGYAEIKITMEQFRLKEDKQAKRQNIHLN  
AAEVEDLSKRLVLDLSIAQVNAVDAVLAASYSKLSGICKDVMVAKKEFTVLESQ  
AAEVESNLALIDQIVKAAVDLTVKPRFOAIEDVEARCTLEKESDFTLSKLDPEVS  
EMLVPMNVDEQGPICLDEFGSNKDTFPGLEEDPGVFATPAGSKCHG"  
complement (23597..26464)  
/gene="At2g06170"  
/note="F5K7.7"  
complement (join(<23597..23821,24081..24417,24700..25436,  
25986..26243,26258..>26464))  
/gene="At2g06170"  
complement (join(23597..23821,24081..24417,24700..25436,  
25986..26243,26258..26464))  
/gene="At2g06170"

Query Match 100.0%; Score 18; DB 8; Length 106716;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTCGGGA 18  
|||||  
Db 49148 TGGCGCCGTTGTCGGGA 49131

RESULT 5  
AP004483 122116 bp DNA linear PLN 14-DEC-2001  
LOCUS AF004483 122116 bp DNA linear PLN 14-DEC-2001  
DEFINITION Lotus japonicus genomic DNA, chromosome 1, clone:LTJ13004, TM0016,

complete sequence.  
ACCESSION AF004483  
VERSION AP004483.1 GI:17736850  
KEYWORDS HMG.  
SOURCE Lotus japonicus DNA, clone:LTJ1 library clone:LTJ13004.  
ORGANISM Lotus japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;  
Lotus.  
REFERENCE 1  
Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. and Tabata, S.  
Structural Analysis of a Lotus japonicus Genome. I. Sequence  
Features and Mapping of Fifty-six TAC clones which cover the 5.4 Mb  
Regions of the Genome  
Unpublished  
2 (bases 1 to 122116)  
Nakamura, Y.  
Direct Submission  
Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research  
Institute, Department of Plant Gene Research: 1532-3, Yana,  
Kisarazu, Chiba 292-0812, Japan (E-mail: ynakam@kazusa.or.jp,  
URL: http://www.kazusa.or.jp, Tel: 81-438-52-3935,  
Fax: 81-438-52-3934)  
FEATURES  
source  
1..122116  
Location/Qualifiers  
/organism="Lotus japonicus"  
/db\_xref="taxon:34305"  
/chromosome="1"  
/clone="LTJ13004"  
/clone\_lib="LTJ library"  
/note="TAC clone:TM0016"

BASE COUNT 37158 a 23562 c 23033 g 38363 t  
ORIGIN

Query Match 100.0%; Score 18; DB 8; Length 122116;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTCGGGA 18  
|||||  
Db 52966 TGGCGCCGTTGTCGGGA 52949

RESULT 6  
AP003054 148246 bp DNA linear PLN 21-MAR-2002  
LOCUS AP003054 148246 bp DNA linear PLN 21-MAR-2002  
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,  
PAC clone: P0436D06.  
ACCESSION AP003054  
VERSION AP003054.2 GI:13359042  
KEYWORDS  
SOURCE  
ORGANISM  
Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,  
clone: P0436D06.  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Echinozoideae; Oryzae; Oryza.  
REFERENCE 1  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Oryza sativa nipponbare (GAS) genomic DNA, chromosome 1, PAC  
clone: P0436D06  
Published Only in Database (2000)  
2 (bases 1 to 148246)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (20-DEC-2000) Takuji Sasaki, National Institute of  
Agrobiological Resources, Rice Genome Research Program, Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasaki@affrc.go.jp, URL: http://rtp.dna.affrc.go.jp/  
On Mar 16, 2001 this sequence version replaced gi:11967926.  
Genes were predicted from the integrated results of the following:

COMMENT

GENSCANL 0. BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI Nonredundant Protein database, and the cDNA sequence database at (<http://ncbi.nlm.nih.gov/blast/db>) and the protein homologies of the coding regions were searched against NCBI Nonredundant Protein database with BLAST2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding dbj accession no. and RGP clone ID. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein. The orientation of the sequence is from 5' to 3' of the PAC clone. This sequence of P0436D06 clone has an overlap with P0507H06 (DBJ:P003144) clone at 5' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

## FEATURES

source

1.148246 "Oryza sativa (japonica cultivar-group)"  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="1"  
 /clone="P0436D06"  
 join(1618..1860,2195..2404)  
 /gene="P0436D06.1"  
 join(1618..1860,2195..2404)  
 /note="P0436D06.1"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAB33265.1"  
 /db\_xref="GI:13359043"  
 /translation="MASPCTCRALKLMVSVDSVTTLTPKRAPTYCPERSISTDLDP  
 AYHLNKKQVITIALAPRPSLRPAVIVPFDNRSAPVAVAKGDKWEAVVQHR  
 RHPSTRATRGSDIWRGSOQYRIKNSMAISLDEFTIDKSTKSG"  
 complement(join(19066..19338,19622..19876,20096..20179,  
 21788..22330,25632..25949))  
 /gene="P0436D06.2"  
 complement(join(19066..19338,19622..19876,20096..20179,  
 21788..22330,25632..25949))  
 /note="hypothetical protein"  
 similar to Arabidopsis thaliana Chromosome 1, F9L1.11"  
 /codon\_start=1  
 /protein\_id="BAB33266.1"  
 /db\_xref="GI:13359044"  
 /translation="MGSSDSQAPLLPRGSHRKEEEEXAAAGKVRCCGDEGGM  
 WREATAEAGRLASLAPMAIVALLQIMQOLISTVMVGLGEVALAGAAINSLTNVSG  
 FSVLWAGLGLTICGAGAEQYHKLALMYRSIILVYVSPVAILIWIPIEVLPL  
 IGOPEISAEVKGKVALMLPGLEAFVAOCLSKFILOSILFPMVLSSTILAFIPL  
 CMFYKRVGMNAGALSVSICWVEYTVGLIVLSPSEKTRAPLTPWEPFSGTST  
 LRLAVSALMLEMWSYELVLGSLIPALFETSVLISGYSVSNELGANGEGA  
 RLAVVALSVILCSAVLVSVTLLALRHFIQIAFSNEEVLINVTVMVPSISTVTD  
 LQGVSGIGVSRGCGWHLGAYVNLGAFIVGVVALFEGFAMHLGGMGFMGMVAVG  
 ATQVTLSTITMTNMRKMGSPFELHKLKLVSAVSKP"  
 join(30567..30579,31283..31494)  
 /gene="P0436D06.3"  
 join(30567..30579,31283..31494)  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAB33267.1"  
 /db\_xref="GI:13359045"  
 /translation="MTDCQKPAIPADPTFTKMTANKTLREFAAPSTDVNAIGPQIMED  
 VDFDKSSLITMAQASPFCKPNEDADAHL"  
 complement(join(32602..32639,32886..32976))  
 /gene="P0436D06.4"  
 complement(join(32602..32639,32886..32976))  
 /note="P0436D06.4"  
 /note="hypothetical protein"

/codon\_start=1  
 /protein\_id="BAB33268.1"  
 /db\_xref="GI:13359046"  
 /translation="NERGSPDGLCFRHLTLRGDQCPPLPAGGVOYGLGFENVV"  
 join(35422..35884,35924..36131,36163..36214)  
 /gene="P0436D06.5"  
 join(35422..35884,35924..36131,36163..36214)  
 /gene="P0436D06.5"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAB33269.1"  
 /db\_xref="GI:13359047"  
 /translation="MAYIMTFNDRKSGRYSHGNDPCPTREAMYNNAAYRPGQO  
 GNNOPPYRGNGNNGNSNPGLKDLVFAQAKTDLASKRLANDKILDNINIKLG  
 FASAPNOISFNKMIETOLALSLVSTNETGRIPGQDSSVEVNAITTTGNTIGS  
 KASSDSADKTEQPERTMDEHCDTRLISFPQSRKRPVSDEQLARVEVIAKHHIV  
 PLLDAMTYSTRDRSQRRSR"  
 complement(36243..37369)  
 /note="3' LTR"  
 38066..38853  
 /note="5' LTR"  
 complement(38856..43616)  
 /gene="P0436D06.6"  
 complement(38856..43616)  
 /gene="P0436D06.6"  
 /note="probably inactive due to stop codon in CDS  
 pseudogene, gag-pol polypeptide"  
 44682..45462  
 /note="3' LTR"  
 join(46344..48451,48465..49341)  
 /gene="P0436D06.7"  
 join(46344..48451,48465..48671,48670..49341)  
 /gene="P0436D06.7"  
 /note="probably inactive due to stop codon in CDS  
 pseudogene, retrovirus-related pol polypeptide from  
 retrotransposon 17.6"  
 /pseudo  
 complement(52237..53379)  
 /note="5' LTR"  
 join(53501..53821,53827..54092,54741..54982,55675..55816,  
 55967..56189)  
 /gene="P0436D06.8"  
 join(53501..53821,53827..54092,54741..54982,55675..55816,  
 55967..56189)  
 /gene="P0436D06.8"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAB33270.1"  
 /db\_xref="GI:13359048"  
 /translation="MLKDVDPKINFVLAPFPMRLQADSSVCYTAGIEDVPVKIRD  
 FFLVDFVDMQTKETPLILRPFSTIGANDVGTGSRFIDKGEVFEQPT  
 EOCSMKVRVSEYSGRSGATODGQVFNHAFPREGNDAOPELDEKTTHTG  
 OEARANSERPALRTKANEVEGKARISITSDGRRRLSRLEFKGSSSHHDES  
 STRSSAVSSEDKARRRLNGLDIDVNRQRAYVILSMVEYHREYRDELKKT  
 GHILVYRELITPLQTEESRAGATRETNMARNEKSSSTPVOISSGVPPSPSE  
 WRSSSRWEIGETTRNMTLDMQTEGLQVLTETIAQMQEQATDQFANINNMHQ  
 QGDDL"  
 join(57860..57947,57977..58160,58266..58488,58712..58743,  
 59094..59199,59234..59612,59620..59747,62210..62476,  
 62757..63233,63954..64091)  
 /gene="P0436D06.9"  
 join(57860..57947,57977..58160,58266..58488,58712..58743,  
 59094..59199,59234..59612,59620..59747,62210..62476,  
 62757..63233,63954..64091)  
 /gene="P0436D06.9"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAB33271.1"  
 /db\_xref="GI:13359049"  
 /translation="MDICRAEORGDCTALOKGTWVSSNMMAKMGCSGHLVLELA  
 IARVYKALFSEGTEDRIPKSGSKILFNPIDQESKILNORPHOIRNGVEGVG  
 IGRGDIIPARCANDQLSAAGTYNELLPFRPRSPVKGCHCTAEFSFYHGASSANDLWS

IHRHIAERAIASGEIFNYPWLSKEMILTKESSPAPPLPEPTLADVSVLEFLIY  
 TKOCHRCIGVARGAVAVLIGFOPSSSSCKRSSLISGNCVAMALCSGSAFMAATLADV  
 VPROTCACVARGAVIDPDAELLRLVSSCOCHAHPCGIDVWMLHSHACVRRRAADAL  
 RVPGNORFLAMTRORGRSAATCADGSHVHERHREKMRSRGKTLPIEPAPV  
 PPTPTMADTLREFPAASAEVNAIGPQINMGDVEDLKSLLITMAQAPFCGKPED  
 ANAHLOFPEICTCTYTKGNFYNGITPMSRDLDAAGAGAFSKVQAGVADILEKVS  
 NMGSEERLOTRGMHTVKEMELFAKLDLMLCKLDHDKRPGCVLADSDVCEV  
 YGAPSSPMTMLVRSTATRTITGARKVRRCICVATTTGTGVHKEVRGCTSHAR  
 TIKELDILANTLYPMPSPWTEKTPQTLLELTTHVHDHTGTLETRYPP  
 complement(65302..65898)  
 /gene="P0436D06.10"  
 complement(65302..65898)  
 /gene="P0436D06.10"  
 complement(65302..65898)  
 /gene="P0436D06.10"  
 /note="hypothetical protein  
 similar to Oryza sativa Chromosome 1, P0463F06.26"  
 /codon\_start=1  
 /protein\_id="BAB33272.1"  
 /db\_xref="GI:13359050"  
 /translation="MCVDETDLNKVCPCDHPFLARIIDQLVDSMACCELLSPDAYSGY  
 HONSMKEDEKTEITPEGVFTYRMPFGISVGTQREIOLGADIGMNVETI  
 DDVYVTKRTGDTLINDLRETNMNRRLKLNKECVSPSGKLLGLVSGGIGIAN  
 PEKIKATENKMSSTRLEKIQRLTGOMALSRFVARGE"  
 complement(join(66849..66972,67147..67435,67625..67669,  
 68423..68608,71933..71987,72830..72928))  
 /gene="P0436D06.11"  
 complement(join(66849..66972,67147..67435,67625..67669,  
 68423..68608,71933..71987,72830..72928))  
 CDS  
 gene  
 Query Match 100.0%; Score 18; DB 8; Length 148246;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 TGGGCGCGTGTGCGGGA 18  
 Db 37423 TGGGCGCGTGTGCGGGA 37440  
 RESULT 7  
 AP003577/c 189744 bp DNA linear HTG 21-MAR-2002  
 LOCUS Oryza sativa (japonica cultivar-group) chromosome 6 clone P0664F03,  
 DEFINITION \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
 ACCESSION AP003577  
 VERSION AP003577.1 GI:13936417  
 KEYWORDS HTG; HTGS PHASE2.  
 SOURCE Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,  
 clone: P0664F03.  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1  
 Sasaki, T., Matsumoto, T. and Yamamoto, K.  
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC  
 clone: P0664F03  
 Published Only in Database (2001)  
 2 (bases 1 to 189744)  
 Sasaki, T., Matsumoto, T. and Yamamoto, K.  
 Direct Submission  
 Submitted (02-MAY-2001) Takuji Sasaki, National Institute of  
 Agrobiological Sciences, Rice Genome Research Program, Kannondai  
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 (E-mail: tsasaki@agr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/  
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
 NOTE: It currently consists of 1 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces is believed  
 to be correct as given, however the sizes of the gaps between them  
 are based on estimates that have provided by the submitter. This  
 sequence will be replaced by the finished sequence as soon as it is  
 available and the accession number will be preserved.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and

\* the accession number will be preserved.  
 FEATURES  
 source  
 1..189744  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="6"  
 /clone="P0664F03"  
 BASE COUNT 54659 a 40138 c 40698 g 53849 t 400 others  
 ORIGIN  
 Query Match 100.0%; Score 18; DB 2; Length 189744;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 TGGGCGCGTGTGCGGGA 18  
 Db 151774 TGGGCGCGTGTGCGGGA 151757  
 RESULT 8  
 AP005160/c 194640 bp DNA linear HTG 17-MAY-2002  
 LOCUS Oryza sativa (japonica cultivar-group) chromosome 7 clone  
 DEFINITION OSJNBa0031C24, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
 ACCESSION AP005160  
 VERSION AP005160.1 GI:20975286  
 KEYWORDS HTG; HTGS PHASE2.  
 SOURCE Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,  
 clone: OSJNBa0031C24.  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1  
 Sasaki, T., Matsumoto, T. and Katayose, Y.  
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC  
 clone: OSJNBa0031C24  
 Published Only in Database (2002)  
 2 (bases 1 to 194640)  
 Sasaki, T., Matsumoto, T. and Katayose, Y.  
 Direct Submission  
 Submitted (15-MAY-2002) Takuji Sasaki, National Institute of  
 Agrobiological Sciences, Rice Genome Research Program, Kannondai  
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 (E-mail: tsasaki@agr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/  
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
 NOTE: It currently consists of 1 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces is believed  
 to be correct as given, however the sizes of the gaps between them  
 are based on estimates that have provided by the submitter. This  
 sequence will be replaced by the finished sequence as soon as it is  
 available and the accession number will be preserved.  
 \* NOTE: This is a 'working draft' sequence.  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 FEATURES  
 source  
 1..194640  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="7"  
 /clone="OSJNBa0031C24"  
 BASE COUNT 57603 a 39166 c 40961 g 56787 t 103 others  
 ORIGIN  
 Query Match 100.0%; Score 18; DB 2; Length 194640;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 TGGGCGCGTGTGCGGGA 18  
 Db 151774 TGGGCGCGTGTGCGGGA 151757



```

gene
  /rpt_type-dispersed
  <29965..>30140
  /gene="Bdi"
  /note="coding region not determined; predicted by BLASTn
  and EST search; barley DCiMA-induced gene; similar to the
  sequence deposited in Genbank Accession Number AJ250665"
  /evidence=experimental
  31684..32211
  /note="miniature inverted transposable element-like
  sequence; type Hormite2"
  /evidence=not_experimental
  /rpt_family="MITE"
  /rpt_type-dispersed
  32253..332962
  /note="incomplete ORF"
  /transposon="Ac-like transposon"
  34560..34564
  /note="5-bp insertion signature (HORPIA-3)"
  34565..36568
  /note="partial; part of polypeptide and 3' untranslated
  or translated region deleted, the LTRs are intact"
  /evidence=not_experimental
  /rpt_family="HORPIA-3"
  /rpt_type-direct
  /transposon="Copia/Y1-like retrotransposon"
  36569..36593
  /note="5-bp insertion signature (HORPIA-3)"
  37785..39765
  /gene="711N16.9"
  join(37785..37972,38394..38577,39727..39765)
  /gene="711N16.9"
  /note="predicted by Genscan; function unknown"
  /product="HV711N16.9"
  /protein_id="AAW22813.1"
  /db_xref="GI:20513852"
  /translation="MPRRPPLSCSVAIRTAARYVPLGRSCMEDPGSAEYMGHDEYK
  LVQFKPAIDCSVORPPPAIVHLMHGVAARYTKRRRSKGDIGARTIDGVVVVVD
  FRNRQASIDSGPPTDLSGYDAVAGCRVRBDGT"
  40068..40310
  /note="Stowaway 3; miniature inverted transposable
  element-like sequence; type Stowaway"
  /rpt_family="MITE"
  /rpt_type-dispersed
  46160..46164
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  46165..47952
  /note="BARE-1 SOL0LTR-1.2"
  /transposon="Copia/Cypsy-like retrotransposon BARE-1"
  47953..47957
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  49087..49120
  /note="Hormites3.1; miniature inverted transposable
  element-like sequence; type Hormite3"
  /evidence=not_experimental
  /rpt_family="MITE"
  /rpt_type-dispersed
  49126..49163
  /note="Hormites3.2; miniature inverted transposable

repeat_region
  /note="miniature inverted transposable element-like
  sequence; type Hormite2"
  /evidence=not_experimental
  /rpt_family="MITE"
  /rpt_type-dispersed
  32253..332962
  /note="incomplete ORF"
  /transposon="Ac-like transposon"
  34560..34564
  /note="5-bp insertion signature (HORPIA-3)"
  34565..36568
  /note="partial; part of polypeptide and 3' untranslated
  or translated region deleted, the LTRs are intact"
  /evidence=not_experimental
  /rpt_family="HORPIA-3"
  /rpt_type-direct
  /transposon="Copia/Y1-like retrotransposon"
  36569..36593
  /note="5-bp insertion signature (HORPIA-3)"
  37785..39765
  /gene="711N16.9"
  join(37785..37972,38394..38577,39727..39765)
  /gene="711N16.9"
  /note="predicted by Genscan; function unknown"
  /product="HV711N16.9"
  /protein_id="AAW22813.1"
  /db_xref="GI:20513852"
  /translation="MPRRPPLSCSVAIRTAARYVPLGRSCMEDPGSAEYMGHDEYK
  LVQFKPAIDCSVORPPPAIVHLMHGVAARYTKRRRSKGDIGARTIDGVVVVVD
  FRNRQASIDSGPPTDLSGYDAVAGCRVRBDGT"
  40068..40310
  /note="Stowaway 3; miniature inverted transposable
  element-like sequence; type Stowaway"
  /rpt_family="MITE"
  /rpt_type-dispersed
  46160..46164
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  46165..47952
  /note="BARE-1 SOL0LTR-1.2"
  /transposon="Copia/Cypsy-like retrotransposon BARE-1"
  47953..47957
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  49087..49120
  /note="Hormites3.1; miniature inverted transposable
  element-like sequence; type Hormite3"
  /evidence=not_experimental
  /rpt_family="MITE"
  /rpt_type-dispersed
  49126..49163
  /note="Hormites3.2; miniature inverted transposable

misc_feature
  /note="5-bp insertion signature (HORPIA-3)"
  34565..36568
  /note="partial; part of polypeptide and 3' untranslated
  or translated region deleted, the LTRs are intact"
  /evidence=not_experimental
  /rpt_family="HORPIA-3"
  /rpt_type-direct
  /transposon="Copia/Y1-like retrotransposon"
  36569..36593
  /note="5-bp insertion signature (HORPIA-3)"
  37785..39765
  /gene="711N16.9"
  join(37785..37972,38394..38577,39727..39765)
  /gene="711N16.9"
  /note="predicted by Genscan; function unknown"
  /product="HV711N16.9"
  /protein_id="AAW22813.1"
  /db_xref="GI:20513852"
  /translation="MPRRPPLSCSVAIRTAARYVPLGRSCMEDPGSAEYMGHDEYK
  LVQFKPAIDCSVORPPPAIVHLMHGVAARYTKRRRSKGDIGARTIDGVVVVVD
  FRNRQASIDSGPPTDLSGYDAVAGCRVRBDGT"
  40068..40310
  /note="Stowaway 3; miniature inverted transposable
  element-like sequence; type Stowaway"
  /rpt_family="MITE"
  /rpt_type-dispersed
  46160..46164
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  46165..47952
  /note="BARE-1 SOL0LTR-1.2"
  /transposon="Copia/Cypsy-like retrotransposon BARE-1"
  47953..47957
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  49087..49120
  /note="Hormites3.1; miniature inverted transposable
  element-like sequence; type Hormite3"
  /evidence=not_experimental
  /rpt_family="MITE"
  /rpt_type-dispersed
  49126..49163
  /note="Hormites3.2; miniature inverted transposable

misc_feature
  /note="5-bp insertion signature (HORPIA-3)"
  34565..36568
  /note="partial; part of polypeptide and 3' untranslated
  or translated region deleted, the LTRs are intact"
  /evidence=not_experimental
  /rpt_family="HORPIA-3"
  /rpt_type-direct
  /transposon="Copia/Y1-like retrotransposon"
  36569..36593
  /note="5-bp insertion signature (HORPIA-3)"
  37785..39765
  /gene="711N16.9"
  join(37785..37972,38394..38577,39727..39765)
  /gene="711N16.9"
  /note="predicted by Genscan; function unknown"
  /product="HV711N16.9"
  /protein_id="AAW22813.1"
  /db_xref="GI:20513852"
  /translation="MPRRPPLSCSVAIRTAARYVPLGRSCMEDPGSAEYMGHDEYK
  LVQFKPAIDCSVORPPPAIVHLMHGVAARYTKRRRSKGDIGARTIDGVVVVVD
  FRNRQASIDSGPPTDLSGYDAVAGCRVRBDGT"
  40068..40310
  /note="Stowaway 3; miniature inverted transposable
  element-like sequence; type Stowaway"
  /rpt_family="MITE"
  /rpt_type-dispersed
  46160..46164
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  46165..47952
  /note="BARE-1 SOL0LTR-1.2"
  /transposon="Copia/Cypsy-like retrotransposon BARE-1"
  47953..47957
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  49087..49120
  /note="Hormites3.1; miniature inverted transposable
  element-like sequence; type Hormite3"
  /evidence=not_experimental
  /rpt_family="MITE"
  /rpt_type-dispersed
  49126..49163
  /note="Hormites3.2; miniature inverted transposable

repeat_region
  /note="miniature inverted transposable element-like
  sequence; type Hormite2"
  /evidence=not_experimental
  /rpt_family="MITE"
  /rpt_type-dispersed
  32253..332962
  /note="incomplete ORF"
  /transposon="Ac-like transposon"
  34560..34564
  /note="5-bp insertion signature (HORPIA-3)"
  34565..36568
  /note="partial; part of polypeptide and 3' untranslated
  or translated region deleted, the LTRs are intact"
  /evidence=not_experimental
  /rpt_family="HORPIA-3"
  /rpt_type-direct
  /transposon="Copia/Y1-like retrotransposon"
  36569..36593
  /note="5-bp insertion signature (HORPIA-3)"
  37785..39765
  /gene="711N16.9"
  join(37785..37972,38394..38577,39727..39765)
  /gene="711N16.9"
  /note="predicted by Genscan; function unknown"
  /product="HV711N16.9"
  /protein_id="AAW22813.1"
  /db_xref="GI:20513852"
  /translation="MPRRPPLSCSVAIRTAARYVPLGRSCMEDPGSAEYMGHDEYK
  LVQFKPAIDCSVORPPPAIVHLMHGVAARYTKRRRSKGDIGARTIDGVVVVVD
  FRNRQASIDSGPPTDLSGYDAVAGCRVRBDGT"
  40068..40310
  /note="Stowaway 3; miniature inverted transposable
  element-like sequence; type Stowaway"
  /rpt_family="MITE"
  /rpt_type-dispersed
  46160..46164
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  46165..47952
  /note="BARE-1 SOL0LTR-1.2"
  /transposon="Copia/Cypsy-like retrotransposon BARE-1"
  47953..47957
  /note="5-bp insertion signature (BARE-1 SOL0LTR-1.2)"
  49087..49120
  /note="Hormites3.1; miniature inverted transposable
  element-like sequence; type Hormite3"
  /evidence=not_experimental
  /rpt_family="MITE"
  /rpt_type-dispersed
  49126..49163
  /note="Hormites3.2; miniature inverted transposable

misc_feature
  /note="5-bp insertion signature (HORPIA-3)"
  34565..36568
  /note="partial; part of polypeptide and 3' untranslated
  or translated region deleted, the LTRs are intact"
  /evidence=not_experimental
  /rpt_family="HORPIA-3"
  /rpt_type-direct
  /transposon="Copia/Y1-like retrotransposon"
  36569..36593
  /note="5-bp insertion signature (HORPIA-3)"
  37785..39765
  /gene="711N16.9"
  join(37785..37972,38394..38577,39727..39765)
  /gene="711N16.9"
  /note="predicted by Genscan; function unknown"
  /product="HV711N16.9"
  /protein_id="AAW22813.1"
  /db_xref="GI:20513852"
  /translation="MPRRPPLSCSVAIRTAARYVPLGRSCMEDPGSAEYMGHDEYK
  LVQFKPAIDCSVORPPPAIVHLMHGVAARYTKRRRSKGDIGARTIDGVVVVVD
  FRNRQASIDSGPPTDLSGYDAVAGCRVRBDGT"
  40068..40310
  /note="Stowaway 3; miniature inverted transposable
  element-like sequence; type Stowaway"
  /rpt_family="MITE"
  /rpt_type-dispersed
  46
```

ACCESSION	AP002458 BA000014
VERSION	AP002458.1 GI:8347610
KEYWORDS	
SOURCE	
ORGANISM	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:TAMU BAC clone:T22C2.
AUTHORS	Arabidopsis thaliana
JOURNAL	Eukaryote: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons; core eudicots; rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE	1 (sites)
TITLE	Nakamura,Y.
REFERENCE	Structural Analysis of Arabidopsis thaliana Chromosome 3. III Unpublished
AUTHORS	2 (bases 1 to 67517)
TITLE	Nakamura,Y.
JOURNAL	Direct Submission Submitted (06-JUN-2000) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yinakamu@kazusa.or.jp, Tel:81-458-52-3935, Fax:81-458-52-3934)
COMMENT	Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/ad_graph.cgi?c=T22C2 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3//), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremlini.zool.lastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is T22B15 and the 3' clone is T10I3.
FEATURES	Location/Qualifiers
source	1..67517
	/organism="Arabidopsis thaliana"
	/strain="Columbia"
	/db_xref="taxon:3702"
	/chromosome="3"
	/clone="T22C2"
	/clone_lib="TAMU BAC"
exon	complement(32..199) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=5
	/evidence-not_experimental
exon	complement(423..758) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=4
	/evidence-not_experimental
exon	complement(909..1318) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=3
	/evidence-not_experimental
exon	complement(1423..1696) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=2
	/evidence-not_experimental
exon	complement(1723..1996) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(2000..2273) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(2300..2573) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(2600..2873) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(2900..3173) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(3200..3473) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(3500..3773) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(3800..4073) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(4100..4373) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(4400..4673) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(4700..4973) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(5000..5273) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(5300..5573) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(5600..5873) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(5900..6173) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(6200..6473) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(6500..6773) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(6800..7073) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(7100..7373) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(7400..7673) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(7700..7973) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"
	/number=1
	/evidence-not_experimental
exon	complement(8000..8273) /product="(+)delta-cadinene synthase (d-cadinene synthase) like" /note="CDS is reported in Acc# AP002062 gene_id:T22B15.28"

gene\_id:"T22B15.28"  
 /number=2  
 /evidence-not\_experimental  
 complement(1840..2094)  
 /product="(+)delta-cadinene synthase (d-cadinene synthase) like"  
 /note="CDS is reported in Acc# AP002062  
 gene\_id:"T22B15.28"  
 /number=1  
 /evidence-not\_experimental  
 complement(join(3258..7313,8240..8346,8681..8834))  
 /note="gene\_id:"T22C2.2"  
 /pseudo  
 /codon\_start=1  
 /evidence-not\_experimental  
 /product="non-LTR retroelement reverse transcriptase-like protein"  
 complement(11186..11911)  
 /note="gene\_id:"T22C2.3"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /product="Tail non-LTR retroelement protein-like"  
 /protein\_id:"BA97081.1"  
 /db\_xref="GI:8777563"  
 /translation="MADNKRRAVODINLGVDDIPPALPEDIYNNAVAENRFTLEGRPV  
 MPRQNLRSIVASMPRTWGSGLVHGRIMEGRQFHTLTLESLETVALDRGFMAENW  
 MLLQWMEPDIPLPPIFWVQIRNGIFPOFNRGVBEHIGALGVLDITDFNVEVAR  
 MGFARVLHMDITHPLRFORHQQFTAGVNTLLRRERYRLRGFCVCGMLTHDFGACLI  
 MNGEERQADDDDDDEEHPTQYHN"  
 complement(join(15703..16226,16332..17328,17506..17657,  
 17792..17980,18269..18867,18953..19231,20205..20875,  
 20974..21359))  
 /note="gene\_id:"T22C2.4"  
 /pseudo  
 /codon\_start=1  
 /evidence-not\_experimental  
 /product="helicase-like protein"  
 26176..28509  
 /note="contains similarity to unknown protein  
 emb|CAB77997.1  
 gene\_id:"T22C2.5"  
 /pseudo  
 /codon\_start=1  
 /evidence-not\_experimental  
 complement(30211..31032)  
 /note="contains similarity to retroelement pol  
 polyprotein  
 gene\_id:"T22C2.6"  
 /pseudo  
 /codon\_start=1  
 /evidence-not\_experimental  
 33467..33844  
 /note="emb|CAB81790.1  
 gene\_id:"T22C2.7  
 similar to unknown protein"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /protein\_id:"BA97082.1"  
 /db\_xref="GI:8777564"  
 /translation="MGADIVADGLTDRPPRLADPAEGLSDRLPEGLSEPADFLVD  
 KLNLRKPHVDLEDFLFTLTKMKTTRHLGAENHTHILFTPKPREKNLSLIGSLAF  
 SKTIKESPLDPCVCLTLEDSLKT"  
 complement(join(34053..35108,35292..35906))  
 /note="contains similarity to non-LTR retroelement reverse  
 transcriptase  
 gene\_id:"T22C2.8"  
 /pseudo  
 /codon\_start=1  
 /evidence-not\_experimental  
 join(35982..36259,36422..37100)  
 /note="gene\_id:"T22C2.9  
 p1r1|T01474  
 similar to unknown protein"

Query Match 94.4% Score 17: DB 8: Length 67517;  
 Best Local Similarity 100.0% Pred. No. 7.2e+02;  
 Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGGCGCGGTGTGCGGG 17  
 Db 43810 TGGCGCGGTGTGCGGG 43826  
 RESULT 11  
 AF028277 537 bp DNA linear PLN 13-APR-1999  
 LOCUS Hordeum chilense RAPD marker IAS-phc9.  
 DEFINITION AF028277  
 VERSION AF028277.1 GI:4581593  
 KEYWORDS  
 SOURCE Hordeum chilense.  
 ORGANISM Hordeum chilense.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

/pseudo  
 /codon\_start=1  
 /evidence-not\_experimental  
 complement(join(40707..41698,41792..42666))  
 /note="gene\_id:"T22C2.10"  
 /pseudo  
 /codon\_start=1  
 /evidence-not\_experimental  
 /product="retroelement pol  
 join(44238..44539,44606..45477,45594..45890,45945..46225)  
 /note="contains similarity to Athalia retroelement ORF1  
 protein  
 gene\_id:"T22C2.11"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /protein\_id:"BA97083.1"  
 /db\_xref="GI:8777565"  
 /translation="MALDPTASNGEFMTKTEARKLLENLAGSNHNHYDVRNRG  
 GSESKQFAELSAKYEQLMRDQKSVNFCEDSSKGVHOFSGDSEDLQAEINFTNGS  
 TNVENPDQOVYPTQAGSQGQKFPQYQONKNYGOQFPQPGATGHASSFQDNEMLTM  
 QQVLEDDKRNADINRVDSMYNDLNGKFATLSSHVKTLENOVSOIVASMRPDTHS  
 GKVKPKGEQCYAIMIOELREIYVAKOVETNNVVEITLVEDEKIVEDEPLSEPPPY  
 VPKLPFGREIORIOREKAYARDEIMOLYVRLPDLVLAHVSYSYLYKYLISNRS  
 IEEGVKLISGEERNAOLVESQROCKEAOITVVEMLAKEMVNCASIPATIPKKIG  
 ITNFKPRISILILADRSYQFPFMGLAENHAHVNFTPTNVEVLELDEKPHDPLNGR  
 PNTNVEAIIIDVRSTINLQIDYALENDIGTRKNPTIEDCIELSDVEEVLDDTKD  
 AHVILPELOVEDIEDIEYKVQCKTSKPSKRSILITSEMKERGRAVVRVGVKVLRMK  
 LTMWRPCEGASSRRRVH"  
 join(47271..47464,47537..47661,47751..47818)  
 /note="gene\_id:"T22C2.12  
 unknown protein"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /protein\_id:"BA97084.1"  
 /db\_xref="GI:8777566"  
 /translation="MSSDNFEFTFASFEDIKSGSLDITNICGVKTYRVLYVYYRNO  
 NDGCPNMEERIEFDIENYEGDKLTCGLAKAYANDFDNNRHCDKILICMRRAKL  
 KVDGLASQGGARKLYESFIESKRS"  
 join(48668..48804,4893..49148,49230..49299,49414..49567,  
 49794..49928,50041..50200,50258..50423,50538..50697,  
 50787..50969)  
 /note="gb|AA06081.1  
 gene\_id:"T22C2.13  
 similar to unknown protein"  
 /codon\_start=1  
 /pseudo  
 /codon\_start=1  
 /evidence-not\_experimental  
 join(51608..51745,51834..52098,52211..52313,52391..52622)  
 /note="gene\_id:"T22C2.14"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /product="replication protein A1-like"

REFERENCE	Pooideae; Triticeae; Hordeum.
AUTHORS	1 (bases 1 to 537)
TITLE	Hernandez,P., Martin,A. and Dorado,G.
JOURNAL	Direct Submission Submitted (06-OCT-1997) Agronomía y Mejora Vegetal, Instituto de Agricultura Sostenible (IAS), Ayda. Menendez Pidal (Alameda del Obispo), Cordoba 14080, Spain
FEATURES	location/Qualifiers
source	1..537
	/organism="Hordeum chilense"
	/strain="H1"
	/db_xref="taxon:15565"
	/clone="IAS-PHCR9"
BASE COUNT	150 a 83 c 111 g 193 t
ORIGIN	
Query Match	91.1%; Score 16.4; DB 8; Length 537;
Best Local Similarity	94.4%; Pred. No. 2.7e+03;
Matches 17; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1 TGCGCCGCGTGTGCGGGA 18 
Db	155 TGCGCCGCGTGTGCGGGA 172
RESULT 12	
LOCUS	AX059280 687 bp DNA linear PAT 17-JAN-2001
DEFINITION	Sequence 13 from Patent WO0055325.
ACCESSION	AX059280
VERSION	AX059280.1 GI:12311385
KEYWORDS	thale cress.
ORGANISM	Arabidopsis thaliana
SOURCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids; II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 1 to 687)
AUTHORS	Preuss,D., Coppenhaver,G. and Kelth,K.
TITLE	Plant chromosome compositions and methods
JOURNAL	Patent: WO 0055325-A 13 21-SEP-2000; The University of Chicago (US)
FEATURES	location/Qualifiers
source	1..687
	/organism="Arabidopsis thaliana"
	/db_xref="taxon:3702"
BASE COUNT	241 a 135 c 115 g 196 t
ORIGIN	
Query Match	91.1%; Score 16.4; DB 6; Length 687;
Best Local Similarity	94.4%; Pred. No. 2.6e+03;
Matches 17; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1 TGCGCCGCGTGTGCGGGA 18 
Db	22 TGCGCCGCGTGTGCGGGA 5
RESULT 13	
LOCUS	BDNAATRNA 1732 bp DNA linear PLN 06-FEB-1997
DEFINITION	B.nigra DNA for tRNA like gene.
ACCESSION	X89901
VERSION	X89901.1 GI:927389
KEYWORDS	Transfer-RNA.
SOURCE	black mustard.
ORGANISM	Brassicata; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids; II; Brassicales; Brassicaceae; Brassica.
REFERENCE	1 (bases 1 to 1732)
AUTHORS	Bhanot,R., Srivastavs,P.S., Delseny,M., This,P., Singh Negi,M. and Lakshminikumaaran,M.

REFERENCE	TITLE	CHARACTERIZATION OF SPECIES-SPECIFIC REPEATS FROM B. NIGRA
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 1732)	
AUTHORS	Lakshmi Kumar, M.S.	
TITLE	Direct Submission	
JOURNAL	Submitted (24-JUL-1995) M.S. Lakshmi Kumar, Tata Energy Research Inst., Barabari Seth Block, Habitat Place, Lodhi Road, New Delhi, 110 003, INDIA	
REFERENCE	3 (bases 1 to 1732)	
AUTHORS	Kapila, R., Negi, M.S., This, P., Delseny, M., Srivastava, P.S. and Lakshmi Kumar, M.	
TITLE	New family of dispersed repeats from <i>Brassica nigra</i> ; characterization and localization	
JOURNAL	Theor. Appl. Genet. 93(9), 1123-1129 (1996)	
FEATURES	Location/Qualifiers	
SOURCE	1..1732	
ORGANISM	" <i>Brassica nigra</i> "	
DB	/db_xref="taxon:3710"	
LOCUS	/clone_lib="lambda zap II"	
DEFINITION	961..1026	
ACCESSION	/note="transfer RNA like molecule"	
VERSION	1268..1732	
KEYWORDS	/note="homology to Bn4 (X67835)"	
SOURCE	BASE COUNT 541 a 406 c 339 g 446 t	
ORGANISM	ORIGIN	
REFERENCE	Query Match	91.1%; Score 16.4; DB 8; Length 1732;
TITLE	Best Local Similarity	94.4%; Pred. No. 2.3e+03;
JOURNAL	Matches 17; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
REFERENCE	QY 1 TGCGCGCGTTGCGGGA 18	
TITLE		
JOURNAL	Db 1029 TGCGCGCGTTGCGGGA 1012	
DEFINITION	Medicago sativa LTR retroelement MCIRE hypothetical protein mRNA,	
ACCESSION	AF439379	4717 bp mRNA linear PLN 02-MAR-2002
VERSION	AF439379	
KEYWORDS	complete cds.	
SOURCE	AF439379.1 GI:19071287	
ORGANISM	Medicago sativa.	
TITLE	Medicago sativa	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales; Medicago.	
AUTHORS	1 (bases 1 to 4717)	
TITLE	Iwashita, S., Naumkina, M., Gau, M., Uchiyama, K., Isobe, S., Mizukami, Y. and Shimamoto, Y.	
JOURNAL	Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa ( <i>Medicago sativa</i> L)	
REFERENCE	Unpublished	
TITLE	2 (bases 1 to 4717)	
AUTHORS	Iwashita, S. and Naumkina, M.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (23-OCT-2001) Graduate School of Agriculture, Hokkaido University, North-9 West-9, Kita-ku, Sapporo, Hokkaido 060-8589, Japan	
FEATURES	Location/Qualifiers	
SOURCE	1..4717	
ORGANISM	" <i>Medicago sativa</i> "	
DB	/db_xref="taxon:3879"	
LOCUS	/clone="pmc197"	
DEFINITION	/transposon="LTR retroelement MCIRE"	
ACCESSION	/note="putative LTR retroelement"	
VERSION	1586..2110	
KEYWORDS	/note="cold-inducible"	
SOURCE	/codon_start=1	
ORGANISM	/evidence="not experimental"	
TITLE	/product="hypothetical protein"	

/protein\_id="AAL84187.1"  
/db\_xref="GI:19071288"

/translation="MINMODPTNQRKLNLSLELIDQATINDNANNOSITIGEOENNEN  
MGRVGEIVEEVFNQNEVTTLECCGAPKDELPOERSDSEDETVDFGEIVKTEKO  
ERLLSKKEIFEQKGRKSKAEIDRVIDEICALFKSLRRTWTSHTLYLKFMELPKRR  
VSKDVLVSSEFMP"

BASE COUNT 1440 a 754 c 892 g 1631 t  
ORIGIN

Query Match 91.1%; Score 16.4; DB 8; Length 4717;  
Best Local Similarity 94.4%; Pred. No. 2e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCGCTGTGCGGGA 18  
|||||  
Db 1261 TGGCGCGCTGTGCGGGA 1278

## RESULT 15

LOCUS SCRI731 6676 bp DNA linear PLN 09-SEP-1992

DEFINITION S.cereale DNA for dispersed repeat sequence (R173-1).

ACCESSION X64100

VERSION X64100.1 GI:21196

KEYWORDS dispersed repeat sequence; R173 family; retrotransposon-like.

SOURCE Secale cereale.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Secale.

1 (bases 1 to 6676)

Langridge, P.

Direct Submission

Submitted (09-JAN-1992) P. Langridge, University of Adelaide,  
Centre for Cereal Biotechnology, Waite Agricultural Res Inst, Glen  
Osmond, South Australia 5064, AUSTRALIA

2 (bases 1 to 6676)

Rogowsky, P.M., Liu, J.Y., Manning, S., Taylor, C. and Langridge, P.

Structural heterogeneity in the R173 family of rye-specific

repetitive DNA sequences

Plant Mol. Biol. 20 (1), 95-102 (1992)

JOURNAL

MEDLINE

PUBMED

1325206

See also X64100-3.

location/Qualifiers

1..6676

/organism="Secale cereale"

/strain="Chinese Spring 1BL/1RS, 1AL/1RS, 1DL/1RS triple

translocation from Rye"

/db\_xref="taxon:4550"

/chromosome="1 (short arm)"

/clone\_lib="lambda EMBL3"

225..1021

/note="long terminal repeat 1"

5325..6121

/note="long terminal repeat 2"

BASE COUNT 1675 a 1374 c 1405 g 2222 t

ORIGIN

Query Match 91.1%; Score 16.4; DB 8; Length 6676;

Best Local Similarity 94.4%; Pred. No. 1.9e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCGCTGTGCGGGA 18

|||||

Db 1031 TGGCGCGCTGTGCGGGA 1048

Search completed: June 20, 2003, 23:23:59

Job time : 78.956 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:46:16 ; Search time 10.3019 Seconds  
(without alignments)  
3934.810 Million cell updates/sec

Title: US-09-965-553-2

Sequence: 1 tggcgccgtgtcgaggga 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq.101002.\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	18	AAZ35254	Plant retroelement
2	18	100.0	9829	AAZ35271	Soybean retroelement
3	18	100.0	10482	AAZ35275	Soybean retroelement
4	18	100.0	12286	AAZ35261	Plant generic retr
5	16.4	91.1	687	AAZ3107	Arabidopsis thalia
6	16.4	91.1	2364	AAH51978	Myobacterium tube
7	16.4	91.1	611590	AAZ22303	Arabidopsis thalia
8	16.4	91.1	4403765	AAI99683	Myobacterium tube
9	16.4	91.1	4403765	AAI99683	Myobacterium tube

10	16.4	91.1	4411529	22	AAI99682	Myobacterium tube
11	16.4	91.1	4411529	22	AAI99682	Myobacterium tube
12	15.4	85.6	509	24	ABO35252	Oligonucleotide fo
13	15.4	85.6	509	24	ABO35253	Oligonucleotide fo
14	15.4	85.6	778	21	AAZ35254	Fusarium venenatum
15	15.4	85.6	1649	24	ABK87940	Human short-chain
16	15.4	85.6	1759	24	ABN59782	Novel human coding
17	15.4	85.6	1770	22	AAH31391	Human secreted pro
18	15.4	85.6	2181	23	ABL15329	Drosophila melanog
19	15.4	85.6	2657	20	AAH84348	Stealth virus nucl
20	15.4	85.6	4165	22	AAH31356	Human secreted pro
21	15.4	85.6	4358	24	AAI67895	Nucleotide sequenc
22	15.4	85.6	7984	24	ABK91610	Modified HIV prote
23	15.4	85.6	9280	24	ABK91615	Modified HIV prote
24	15.4	85.6	9285	24	ABK91608	Modified HIV prote
25	15.4	85.6	11765	23	ABL15328	Drosophila melanog
26	15	83.3	3510	24	ABN79856	Fungal ZBC gene se
27	14.8	82.2	27	21	AAZ35424	Plant retroelement
28	14.8	82.2	128	22	ABA76135	Human foetal liver
29	14.8	82.2	128	22	ABA40687	Probe #19153 for g
30	14.8	82.2	128	22	AAK24801	Human brain expres
31	14.8	82.2	128	22	AAK50798	Human bone marrow
32	14.8	82.2	128	22	AAI27820	Probe #17753 for g
33	14.8	82.2	128	22	AAI56792	Human genome-deriv
34	14.8	82.2	128	24	ABS24287	Human cytoskeletal
35	14.8	82.2	466	22	AAZ29778	DNA encoding novel
36	14.8	82.2	466	23	ABK43860	Human ORE3342 cDNA
37	14.8	82.2	510	24	ABN78395	Oligonucleotide fo
38	14.8	82.2	528	24	ABO29034	Oligonucleotide fo
39	14.8	82.2	528	24	ABO29035	Oligonucleotide fo
40	14.8	82.2	561	22	ABA63912	Human foetal liver
41	14.8	82.2	561	22	ABA31089	Probe #9555 for ge
42	14.8	82.2	561	22	AAK12418	Human brain expres
43	14.8	82.2	561	22	AAK38131	Human bone marrow
44	14.8	82.2	561	22	AAI18911	Probe #8844 for ge
45	14.8	82.2	561	22	AAI44032	Probe #12718 used

## ALIGNMENTS

RESULT 1	AAZ35254	standard; DNA; 18 BP.
AAZ35254	AAZ35254	
AC	AAZ35254	
XX		
DT	27-MAR-2000	(first entry)
XX		
DE	Plant retroelement primer binding site version 2.	
XX		
KW	Retroelement; retrovirus; transgenic plant; gene transfer;	
KW	primer binding site; soybean; ss.	
XX		
OS	Glycine max.	
XX		
PN	WO9960842-A2.	
XX		
PD	02-DEC-1999.	
XX		
PF	28-MAY-1999;	99WO-US11858.
XX		
PR	29-MAY-1998;	98US-0087125.
XX	28-MAY-1999;	99US-0087125.
XX		
PA	(WRIGHT) WRIGHT D A.	
PA	(VOYT/) VOYTAS D F.	
XX		
PI	Wright DA, Voytas DF;	
XX		
DR	WPI; 2000-105586/09.	
XX		
PT	New nucleic acid molecules for imparting agronomically significant	

PT characters to plants, especially soybean  
 XX  
 PS Claim 1(a); Page 72; 118bp; English.  
 XX  
 CC This oligonucleotide represents a soybean retroelement primer  
 CC binding site (version 2). The invention provides molecular tools  
 CC in the form of retroelements and retroelement-containing vectors,  
 CC cells and plants. Methods are provided for introducing the  
 CC retroelement into cells, especially when the retroelement carries  
 CC at least 1 agronomically-significant characteristic. In a  
 CC preferred method, a helper cell line which expresses gag, pol and  
 CC env sequences is used to enable transfer of a secondary construct  
 CC which carries an agronomically-significant characteristic and has  
 CC retroelement sequences that allow for replication and integration.  
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
 CC sequence selected from a retroelement primer binding site,  
 CC envelope, gag, integrase, reverse transcriptase, protease or  
 CC RNase-H sequence (see AA235274-61). Also provided are plant  
 CC retroviral particles that are used to transfer the nucleic acids  
 CC into plant cells.  
 CC  
 SO Sequence 18 BP; 1 A; 4 C; 9 G; 4 T; 0 other:  
 Query Match 100.0%; Score 18; DB 21; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TGCGCGCGTGTGCGGA 18  
 DB 1 TGCGCGCGTGTGCGGA 18  
 RESULT 2  
 AA235271  
 ID AA235271 standard; DNA: 9829 BP.  
 AC AA235271;  
 XX  
 DT 27-MAR-2000 (first entry)  
 XX  
 DE Soybean retroelement Calypso 1-1.  
 XX  
 KW Retroelement; retrovirus; transgenic plant; gene transfer;  
 KM Calypso 1-1; soybean; ss.  
 XX  
 OS Glycine max.  
 XX  
 PN WO960842-A2.  
 XX  
 PD 02-DEC-1999.  
 XX  
 PF 28-MAY-1999; 99WO-US11858.  
 XX  
 PR 29-MAY-1998; 98US-0087125.  
 XX  
 PR 28-MAY-1999; 99US-0087125.  
 XX  
 PA (WRIG/) WRIGHT D A.  
 PA (VOYT/) VOYTAS D F.  
 PT Wright DA, Voytas DF;  
 XX  
 PI Wright DA, Voytas DF;  
 XX  
 DR WPI; 2000-105586/09.  
 XX  
 PT New nucleic acid molecules for imparting agronomically significant  
 PT characters to plants, especially soybean  
 XX  
 PS Example 3; Page 95-98; 118bp; English.  
 XX  
 CC This is the nucleotide sequence of the Calypso 1-1 retroelement of  
 CC soybean. It was identified by screening of a soybean lambda  
 CC library using a reverse transcriptase probe. 2 groups of soybean  
 CC retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3  
 CC (see AA235271-73) and Calypso 2-1 and 2-2 (see AA235274-75). The

CC retroelements include gag, pol, env and primer binding site  
 CC sequences that can be used in constructs of the invention. The  
 CC invention provides molecular tools in the form of retroelements and  
 CC retroelement-containing vectors, cells and plants. Methods are  
 CC provided for introducing the retroelements into cells, especially  
 CC when the retroelement carries at least 1 agronomically-significant  
 CC characteristic (ACS). In a preferred method, a helper cell line  
 CC which expresses gag, pol and env sequences is used to enable  
 CC transfer of a secondary construct which carries an ACS and has  
 CC retroelement sequences that allow for replication and integration.  
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
 CC sequence selected from a retroelement primer binding site, envelope,  
 CC gag, integrase, reverse transcriptase, protease or RNase-H sequence  
 CC (see AA235274-61). Also provided are plant retroviral particles that  
 CC particles are used to transfer the nucleic acids into plant cells.  
 CC  
 SO Sequence 9829 BP; 2948 A; 2067 C; 2260 G; 2554 T; 0 other:  
 Query Match 100.0%; Score 18; DB 21; Length 9829;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TGCGCGCGTGTGCGGA 18  
 DB 1312 TGCGCGCGTGTGCGGA 1329  
 RESULT 3  
 AA235275  
 ID AA235275 standard; DNA: 10482 BP.  
 AC AA235275;  
 XX  
 DT 27-MAR-2000 (first entry)  
 XX  
 DE Soybean retroelement Calypso 2-2.  
 XX  
 KW Retroelement; retrovirus; transgenic plant; gene transfer;  
 KM Calypso 2-2; soybean; ss.  
 XX  
 OS Glycine max.  
 XX  
 PN WO960842-A2.  
 XX  
 PD 02-DEC-1999.  
 XX  
 PF 28-MAY-1999; 99WO-US11858.  
 XX  
 PR 29-MAY-1998; 98US-0087125.  
 XX  
 PR 28-MAY-1999; 99US-0087125.  
 XX  
 PA (WRIG/) WRIGHT D A.  
 PA (VOYT/) VOYTAS D F.  
 PT Wright DA, Voytas DF;  
 XX  
 PI Wright DA, Voytas DF;  
 XX  
 DR WPI; 2000-105586/09.  
 XX  
 PT New nucleic acid molecules for imparting agronomically significant  
 PT characters to plants, especially soybean  
 XX  
 PS Example 3; Page 107-111; 118bp; English.  
 XX  
 CC This is the nucleotide sequence of the Calypso 2-2 retroelement of  
 CC soybean. It was identified by screening of a soybean lambda  
 CC library using a reverse transcriptase probe. 2 groups of soybean  
 CC retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3  
 CC (see AA235271-73) and Calypso 2-1 and 2-2 (see AA235274-75). The  
 CC retroelements include gag, pol, env and primer binding site  
 CC sequences that can be used in constructs of the invention. The  
 CC invention provides molecular tools in the form of retroelements and  
 CC retroelement-containing vectors, cells and plants. Methods are  
 CC provided for introducing the retroelements into cells, especially

CC when the retroelement carries at least 1 agronomically-significant  
 CC characteristic (ACS). In a preferred method, a helper cell line  
 CC which expresses gag, pol and env sequences is used to enable  
 CC transfer of a secondary construct which carries an ASC and has  
 CC retroelement sequences that allow for replication and integration.  
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
 CC sequence selected from a retroelement primer binding site, envelope,  
 CC gag, integrase, reverse transcriptase, protease or RNase-H sequence  
 CC (see AA235254-61). Also provided are plant retroviral particles that  
 CC particles are used to transfer the nucleic acids into plant cells.

XX Sequence 10482 BP; 3181 A; 1842 C; 2297 G; 3162 T; 0 other;

Query Match 100.0%; Score 18; DB 21; Length 10482;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTCGGGGA 18

DB 1448 TGGCGCCGTTGTCGGGGA 1465

RESULT 4  
 AA235261

ID AA235261 standard; DNA; 12286 BP.

XX AA235261;

XX 27-MAR-2000 (first entry)

XX Plant generic retroelement.

XX Retroelement; retrovirus; transgenic plant; gene transfer;

XX soybean; pea; calypso; athila; cyclops; ss.

XX Glycine max.

XX Arabidopsis thaliana.

XX Pleum sativum.

XX Key Location/Qualifiers

XX CDS 1482..6887

XX MO9960842-A2.

XX 02-DEC-1999.

XX 28-MAY-1999; 99WO-US11858.

XX 29-MAY-1998; 98US-0087125.

XX 28-MAY-1999; 99US-0087125.

XX (WRIG/) WRIGHT D A.

XX (VOYT/) VOYTAS D F.

XX Wright DA, Voytas DF;

XX WPI; 2000-105586/09.

XX P-PSDB; AAY32434.

XX New nucleic acid molecules for imparting agronomically significant

XX characters to plants, especially soybean

XX Claim 1(n): Page 84-88; 118pp; English.

XX The present sequence comprises a generic plant retroelement

XX obtained from retrovirus-like elements (retroelements) calypso of

XX soybean, cyclops of pea and athila of Arabidopsis thaliana. The

XX invention provides molecular tools in the form of retroelements and

XX retroelement-containing vectors, cells and plants. Methods are

XX provided for introducing the retroelements into cells, especially

XX when the retroelement carries at least 1 agronomically-significant

XX characteristic (ACS). In a preferred method, a helper cell line

CC which expresses gag, pol and env sequences is used to enable  
 CC transfer of a secondary construct which carries an ASC and has  
 CC retroelement sequences that allow for replication and integration.  
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid  
 CC sequence selected from a retroelement primer binding site,  
 CC envelope, gag, integrase, reverse transcriptase, protease or  
 CC RNase-H sequence (see AA235254-61). Also provided are plant  
 CC retroviral particles that are used to transfer the nucleic acids  
 CC into plant cells.

XX Sequence 12286 BP; 3748 A; 2540 C; 2767 G; 3231 T; 0 other;

Query Match 100.0%; Score 18; DB 21; Length 12286;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTCGGGGA 18

DB 1312 TGGCGCCGTTGTCGGGGA 1329

RESULT 5  
 AAF22107/c

ID AAF22107 standard; DNA; 687 BP.

XX AAF22107;

XX 20-MAR-2001 (first entry)

XX Arabidopsis thaliana centromere conserved sequence 1-2 #13.

XX Centromere; michrosome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Claim 108; Page 274; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 687 BP; 241 A; 135 C; 115 G; 196 T; 0 other;

Query Match 91.1%; Score 16.4; DB 21; Length 687;

Best Local Similarity 94.4%; Pred. No. 1.6e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTCGGGGA 18

```
Db          22 TGGCGCGCTGCGGGA 5
|||||
RESULT 6
AAH51978/C
ID AAH51978 standard; DNA: 2364 BP.
XX
AC AAH51978;
XX
DT 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 32.
XX
KM Drug target; growth; organism viability; characterisation; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200135317-A1.
XX
PD 17-MAY-2001.
XX
PF 13-NOV-2000; 2000WO-US31152.
XX
PR 12-NOV-1999; 99US-0165086.
PR 12-NOV-1999; 99US-0165124.
PR 01-FEB-2000; 2000US-0179531.
XX
PA (REGC ) UNIV CALIFORNIA.
PI Eisenberg D, Rotstein SH, Marcotte EM;
XX
DR WPI: 2001-329193/34.
DR P-PSDB; AAC61127.
XX
XX
PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences
XX
PS Disclosure; Page 71-72; 207pp; English.
XX
XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAC61096 - AAC61241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analyzing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterizing the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
XX
SQ Sequence 2364 BP; 467 A; 775 C; 715 G; 407 T; 0 other;
XX
Query Match          91.1%; Score 16.4; DB 22; Length 2364;
Best Local Similarity 94.4%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY      1 TGGCGCGCTGTCGGGCA 18
      |||||
Db      1084 TGGCGCGCTGTCGGGCA 1067
XX
RESULT 7
AAF22303
ID AAF22303 standard; DNA: 611590 BP.
XX
AC AAF22303;
XX
```

```
DT 20-MAR-2001 (first entry)
XX
XX Arabidopsis thaliana chromosome 2 centromere.
XX
XX Centromere; microsome; vector; ds.
XX
XX Arabidopsis thaliana.
XX
XX WO200055325-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-US07392.
XX
PR 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
XX (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Kelch K;
XX
DR WPI: 2000-587529/55.
XX
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
PS Claim 45; Page 820-959; 1449pp; English.
XX
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 184983 T; 0 other;
XX
Query Match          91.1%; Score 16.4; DB 21; Length 611590;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY      1 TGGCGCGCTGTCGGGCA 18
      |||||
Db      100117 TGGCGACGTGTCGGGCA 100134
XX
RESULT 8
AAI99683
ID AAI99683 standard; DNA: 4403765 BP.
XX
AC AAI99683;
XX
DT 15-JAN-2002 (first entry)
XX
XX
DE Mycobacterium tuberculosis strain H37RV genome SEQ ID NO 2.
XX
KM Mycobacterium tuberculosis; strain H37RV; strain CDC 1551; genome;
XX variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
```

PA (GENO-) INST GENOMIC RES.  
XX  
PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX  
XX WPI: 2001-647261/74.  
XX  
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
PT determining the nucleotide sequence of the strain at positions in the  
PT genome corresponding to positions where M. tuberculosis strains CDC  
PT 1551 and H37Rv differ  
XX  
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
XX  
CC The invention relates to evaluating strain variation within and between  
CC different populations of the tuberculosis bacterial pathogen.  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
CC H37Rv (AA199682). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.  
XX  
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;  
XX  
Query Match 91.1%; Score 16.4; DB 22; Length 4403765;  
Best Local Similarity 94.4%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
QY 1 TGGCGCCGCTGTGCGGGA 18  
DB 4130872 TGGCGCCGCTGTGCGGGA 4130889  
XX  
RESULT 9  
AA199683/C  
ID AA199683 standard; DNA: 4403765 BP.  
XX  
AC AA199683;  
XX  
DT 15-JAN-2002 (first entry)  
XX  
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
XX  
KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
XX US6294328-B1.  
XX  
XX 25-SEP-2001.  
XX  
XX 24-JUN-1998; 98US-0103840.  
XX  
XX 24-JUN-1998; 98US-0103840.  
XX  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX  
XX WPI: 2001-647261/74.  
XX  
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
PT determining the nucleotide sequence of the strain at positions in the  
PT genome corresponding to positions where M. tuberculosis strains CDC  
PT 1551 and H37Rv differ  
XX  
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

XX  
CC The invention relates to evaluating strain variation within and between  
CC different populations of the tuberculosis bacterial pathogen.  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
CC H37Rv (AA199682). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.  
XX  
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;  
XX  
Query Match 91.1%; Score 16.4; DB 22; Length 4403765;  
Best Local Similarity 94.4%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
QY 1 TGGCGCCGCTGTGCGGGA 18  
DB 809593 TGGCGCCGCTGTGCGGGA 809576  
XX  
RESULT 10  
AA199682  
ID AA199682 standard; DNA: 4411529 BP.  
XX  
AC AA199682;  
XX  
DT 15-JAN-2002 (first entry)  
XX  
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.  
XX  
KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
XX US6294328-B1.  
XX  
XX 25-SEP-2001.  
XX  
XX 24-JUN-1998; 98US-0103840.  
XX  
XX 24-JUN-1998; 98US-0103840.  
XX  
XX 24-JUN-1998; 98US-0103840.  
XX  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX  
XX WPI: 2001-647261/74.  
XX  
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
PT determining the nucleotide sequence of the strain at positions in the  
PT genome corresponding to positions where M. tuberculosis strains CDC  
PT 1551 and H37Rv differ  
XX  
XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.  
XX  
CC The invention relates to evaluating strain variation within and between  
CC different populations of the tuberculosis bacterial pathogen.  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
CC H37Rv (AA199682). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.  
XX  
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;  
Query Match 91.1%; Score 16.4; DB 22; Length 4411529;  
Best Local Similarity 94.4%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 TGGCGCCGTTGTCGGGGA 18  
DB 4138678 TGGCGCCGTTGTCGGGGA 4138695  
RESULT 11  
AA199682/C  
ID AA199682 standard; DNA; 4411529 BP.  
XX  
AC AA199682;  
XX  
DT 15-JAN-2002 (first entry)  
XX  
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.  
XX  
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN US6294328-B1.  
XX  
PD 25-SEP-2001.  
XX  
PE 24-JUN-1998; 98US-0103840.  
XX  
PR 24-JUN-1998; 98US-0103840.  
XX  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX  
DR WPI: 2001-647261/74.  
XX  
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
PT determining the nucleotide sequence of the strain at positions in the  
PT genome corresponding to positions where M. tuberculosis strains CDC  
PT 1551 and H37Rv differ.  
XX  
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.  
XX  
CC The invention relates to evaluating strain variation within and between  
CC different populations of the tuberculosis bacterial pathogen,  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
CC H37Rv (AA199682). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.  
XX  
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;  
Query Match 91.1%; Score 16.4; DB 22; Length 4411529;  
Best Local Similarity 94.4%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 TGGCGCCGTTGTCGGGGA 18  
DB 807416 TGGCGCCGTTGTCGGGGA 807399

RESULT 12  
ABQ35252  
ID ABQ35252 standard; DNA; 509 BP.  
XX  
AC ABQ35252;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21843.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PE 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guelzig D;  
XX  
DR WPI: 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA.  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 509 BP; 49 A; 63 C; 174 G; 223 T; 0 other;  
Query Match 85.6%; Score 15.4; DB 24; Length 509;  
Best Local Similarity 94.1%; Pred. No. 4.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 2 GCGCGCGTTGTCGGGGA 18  
DB 207 GCGCGCGTTGTCGGGGA 223  
RESULT 13

ABQ35253/c  
 ID ABQ35253 standard; DNA; 509 BP.  
 AC ABQ35253;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21844.  
 XX  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218662-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EPI0074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIC-) EPIDENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 DR WPI: 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 XX  
 PS Claim 12: 56bp + Sequence Listing: 56bp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ3410-ABQ3412 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 509 BP; 223 A; 174 C; 63 G; 49 T; 0 other;

Query Match 85.6%; Score 15.4; DB 24; Length 509;  
 Best Local Similarity 94.1%; Pred. No. 4.8e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGGCGCCGTTGCGGGA 18  
 ||||| ||||| ||||| |||||  
 Db 303 GGGCGCGTTGTCGGGA 287

RESULT 14  
 AAF08402/c  
 ID AAF08402 standard; CDNA; 778 BP.  
 XX

AC AAF08402;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Fusarium venenatum EST SEQ ID NO:925.  
 XX  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Fusarium venenatum.  
 XX  
 PN WO200056762-A2.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 22-MAR-2000; 2000WO-US07781.  
 XX  
 PR 22-MAR-1999; 99US-0273623.  
 XX  
 PA (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX  
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 DR WPI: 2000-594572/56.  
 XX  
 PT Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX  
 PS Claim 86; Page 737; 3161pp; English.  
 XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
 CC niger; AAF1854 to AAF1878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.  
 XX  
 SQ Sequence 778 BP; 184 A; 232 C; 184 G; 176 T; 2 other;

Query Match 85.6%; Score 15.4; DB 21; Length 778;  
 Best Local Similarity 94.1%; Pred. No. 4.7e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGTCGGG 17  
 ||||| ||||| ||||| |||||  
 Db 523 TGGCGCCGTTGTCGGG 507

RESULT 15  
 AAF087940

ID ABR87940 standard; cDNA; 1649 BP.  
XX ABR87940;  
AC  
XX  
DT 07-OCT-2002 (first entry)  
XX  
XX  
DE Human short-chain dehydrogenase/reductase, 25206, cDNA.  
XX  
XX  
KM Human; gene: ss; cytosolic; noctropic; neuroprotective; antitumour;  
KM 25206; short-chain dehydrogenase; short-chain reductase; SDR; retinol;  
KM retinal; retinoic acid; embryonic development; spermatogenesis;  
KM epithelial differentiation; metabolism; development; growth;  
KM proliferative disorder; haematopoietic disorder; cancer; tumour;  
KM differentiative disorder; carcinoma; sarcoma; Hodgkin's disease;  
KM neurodegenerative disorder; Alzheimer's disease; Parkinsonism;  
KM progressive supranuclear palsy; reproductive disorder; ovarian tumours;  
KM polycystic ovarian disease.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 213..1073  
FT /tag- a  
FT /product- "25206"  
FT /note- "This CDS is specifically claimed in claim 1"  
XX  
XX  
PN MO200244356-A2.  
XX  
PD 06-JUN-2002.  
XX  
XX  
PD 29-NOV-2001; 2001MO-US45040.  
XX  
XX  
PF 30-NOV-2000; 2000US-250186P.  
XX  
XX  
PR (MILL-) MILLENIUM PHARM INC.  
XX  
XX  
PI Meyers RE, Macbeth KJ;  
XX  
XX  
DR WPI: 2002-547697/58.  
DR P-PSDB; AAU99344.  
XX  
XX  
PT Novel isolated 25206 polypeptide, a human short-chain  
PT dehydrogenase/reductase family member, useful as diagnostic targets and  
PT therapeutic agents for treating cancer, Alzheimer's disease,  
PT Parkinsonism, ovarian tumors  
XX  
XX  
PS Claim 1: Page 101-102; 117pp; English.  
XX  
XX  
CC The invention discloses an isolated 25206 polypeptide, a human  
CC short-chain dehydrogenase/reductase (SDR) family member. Members of this  
CC family catalyse the reversible, rate limiting conversion of retinol to  
CC retinal. Retinal is then converted to retinoic acid which plays a key  
CC role in the regulation of embryonic development, spermatogenesis and  
CC epithelial differentiation. Short chain dehydrogenases are important in  
CC the metabolism of small molecules, production/removal of biologically  
CC important molecules that modulate development and growth, elimination of  
CC toxins and associated physiological processes and pathological  
CC conditions. The polynucleotide and polypeptide are useful for identifying  
CC compounds which binds to them, for identifying compounds which modulate  
CC the activity of the polypeptide, raising antibodies, for detecting the  
CC presence of the polypeptide in a sample, for modulating aberrant activity  
CC of 25206-expressing cells (e.g. a cancerous, pre-cancerous or neural  
CC cell) and for treating or preventing disorders characterised by the  
CC aberrant activity. The 25206 molecules can act as diagnostic targets and  
CC therapeutic agents for controlling one or more cellular proliferative,  
CC including haematopoietic disorders, or differentiative disorders (e.g.  
CC cancer, tumours, carcinoma, sarcoma and Hodgkin's disease), neural (e.g.  
CC neurodegenerative disorders including Alzheimer's disease, Parkinsonism  
CC and progressive supranuclear palsy) and reproductive disorders (e.g.  
CC polycystic ovarian disease and ovarian tumours). The sequence presented  
CC is the human short-chain dehydrogenase/reductase, 25206, cDNA.  
XX  
XX  
SQ Sequence 1649 BP; 275 A; 543 C; 532 G; 299 T; 0 other;

Query Match: 85.6%; Score 15.4; DB 24; Length 1649;  
Best Local Similarity 94.1%; Pred. NO. 4.6e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GCGCCGTTGTGCGGGA 18  
||||| |||||||  
DB 1274 GCGCCCTTGTGCGGGA 1290

Search completed: June 20, 2003, 21:58:31  
Job time : 22.3019.secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:44:01 ; Search time 2.17925 Seconds  
(without alignments)  
2533.070 Million cell updates/sec

Title: US-09-965-553-2  
Perfect score: 18  
Sequence: 1 tggcgcgctgtcgggga 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCROS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	18	100.0	18	4	US-09-322-478-2		Sequence 2, Appl1
	2	18	100.0	27	4	US-09-322-478-39		Sequence 39, Appl1
	3	18	100.0	9829	4	US-09-322-478-19		Sequence 19, Appl1
	4	18	100.0	10482	4	US-09-322-478-23		Sequence 23, Appl1
	5	18	100.0	12286	4	US-09-322-478-17		Sequence 17, Appl1
	6	16.4	91.1	4403765	4	US-09-103-840A-2		Sequence 2, Appl1
	7	16.4	91.1	4403765	4	US-09-103-840A-2		Sequence 2, Appl1
	8	16.4	91.1	4411529	4	US-09-103-840A-1		Sequence 1, Appl1
	9	16.4	91.1	4411529	4	US-09-103-840A-1		Sequence 1, Appl1
	10	14.8	82.2	1018	4	US-09-452-239-35		Sequence 35, Appl1
	11	14.8	82.2	1078	4	US-09-452-239-41		Sequence 41, Appl1
	12	14.8	82.2	2409	4	US-09-230-225B-3		Sequence 3, Appl1
	13	14.8	82.2	3348	4	US-09-302-620B-94		Sequence 94, Appl1
	14	14.8	82.2	6909	4	US-09-199-637A-111		Sequence 111, Appl1
	15	14.4	80.0	231	2	US-08-611-757-14		Sequence 14, Appl1
	16	14.4	80.0	231	5	US-08-611-757-14		Sequence 14, Appl1
	17	14.4	80.0	1851	4	US-08-867-611-29		Sequence 29, Appl1
	18	14.4	80.0	1851	4	US-08-867-611-29		Sequence 29, Appl1
	19	13.8	76.7	41	1	US-08-530-492-143		Sequence 143, Appl1
	20	13.8	76.7	41	1	US-08-530-492-143		Sequence 143, Appl1
	21	13.8	76.7	273	4	US-09-068-101-3		Sequence 3, Appl1
	22	13.8	76.7	462	4	US-09-199-637A-212		Sequence 212, Appl1
	23	13.8	76.7	624	4	US-09-068-101-6		Sequence 6, Appl1
	24	13.8	76.7	624	4	US-09-199-637A-210		Sequence 210, Appl1
	25	13.8	76.7	750	3	US-09-167-717-3		Sequence 3, Appl1
	26	13.8	76.7	1271	4	US-09-231-227-1		Sequence 1, Appl1
	27	13.8	76.7	1386	1	US-08-672-571A-4		Sequence 4, Appl1

C	28	13.8	76.7	1413	4	US-08-984-709A-52	Sequence 52, Appl1
C	29	13.8	76.7	1425	2	US-08-356-060A-6	Sequence 6, Appl1
C	30	13.8	76.7	1425	4	US-08-460-900C-6	Sequence 6, Appl1
C	31	13.8	76.7	1425	4	US-08-674-509B-6	Sequence 6, Appl1
C	32	13.8	76.7	1425	4	US-08-954-698-6	Sequence 6, Appl1
C	33	13.8	76.7	1425	4	US-08-957-874-6	Sequence 6, Appl1
C	34	13.8	76.7	1425	4	US-09-325-256-10	Sequence 10, Appl1
C	35	13.8	76.7	1473	1	US-08-672-571A-2	Sequence 2, Appl1
C	36	13.8	76.7	1576	1	US-08-748-591-5	Sequence 5, Appl1
C	37	13.8	76.7	1576	1	US-08-748-591-10	Sequence 10, Appl1
C	38	13.8	76.7	1669	4	US-08-984-709A-51	Sequence 51, Appl1
C	39	13.8	76.7	1931	1	US-08-530-492-2	Sequence 2, Appl1
C	40	13.8	76.7	1931	4	US-08-906-517-2	Sequence 2, Appl1
C	41	13.8	76.7	2081	4	US-09-149-476-123	Sequence 123, Appl1
C	42	13.8	76.7	2414	6	5248599-1	Patent No. 5248599
C	43	13.8	76.7	2436	4	US-09-199-637A-92	Sequence 92, Appl1
C	44	13.8	76.7	2572	4	US-09-221-017B-225	Sequence 225, Appl1
C	45	13.8	76.7	2804	1	US-08-446-794A-3	Sequence 3, Appl1

#### ALIGNMENTS

```
RESULT 1
US-09-322-478-2
; Sequence 2, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT FILING DATE: 1999-05-28
; EARLIER FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-2

Query Match      100.0%; Score 18; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCCGTTGTCGGGGA 18
        |||||
DB      1 TGGCGCCGTTGTCGGGGA 18

RESULT 2
US-09-322-478-39
; Sequence 39, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT FILING DATE: 1999-05-28
; EARLIER FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 27
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-39
```

Query Match 100.0%; Score 18; DB 4; Length 27;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18  
|||||  
Db 5 TGGCGCCGTTGTGCGGGA 22

RESULT 3

US-09-322-478-19  
; Sequence 19, Application US/09322478  
; Patent No. 6331662  
; GENERAL INFORMATION:

; APPLICANT: Wright, David A.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/322,478  
; EARLIER FILING DATE: 1999-05-28  
; EARLIER APPLICATION NUMBER: 60/087125  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 9829  
; TYPE: DNA  
; ORGANISM: Glycine max

US-09-322-478-19

Query Match 100.0%; Score 18; DB 4; Length 9829;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18  
|||||  
Db 1312 TGGCGCCGTTGTGCGGGA 1329

RESULT 4

US-09-322-478-23  
; Sequence 23, Application US/09322478  
; Patent No. 6331662  
; GENERAL INFORMATION:

; APPLICANT: Wright, David A.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/322,478  
; EARLIER FILING DATE: 1999-05-28  
; EARLIER APPLICATION NUMBER: 60/087125  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 10482  
; TYPE: DNA  
; ORGANISM: Glycine max

Query Match 100.0%; Score 18; DB 4; Length 10482;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18  
|||||  
Db 1448 TGGCGCCGTTGTGCGGGA 1465

RESULT 5  
US-09-322-478-17  
; Sequence 17, Application US/09322478

; Patent No. 6331662  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/322,478  
; EARLIER FILING DATE: 1999-05-28  
; EARLIER APPLICATION NUMBER: 60/087125  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 12286  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: plant  
US-09-322-478-17

Query Match 100.0%; Score 18; DB 4; Length 12286;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18  
|||||  
Db 1312 TGGCGCCGTTGTGCGGGA 1329

RESULT 6

US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-2007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 91.1%; Score 16.4; DB 4; Length 4403765;  
Best Local Similarity 94.4%; Pred. No. 18;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18  
|||||  
Db 4130872 TGGCGCCGTTGTGCGGGA 4130889

RESULT 7

US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.

Query Match	82.2%	Score 14.8	DB 4	Length 1018
Best Local Similarity	88.9%	Pred. No. 1.5e+02		
Matches 16	Conservative	0	Mismatches 2	Indels 0
				Gaps 0
QY	1	TTGCCCCCTTGTGCGGGA	18	



RESULT 14  
US-09-199-637A-111  
Sequence 111, Application US/09199637A  
Patent No. 6355411  
GENERAL INFORMATION:  
APPLICANT: Ausubel, Frederick  
APPLICANT: Goodman, Howard M.  
APPLICANT: Rahne, Laurence G.  
APPLICANT: Mahajan-Miklos, Shaila  
APPLICANT: Cao, Hui  
APPLICANT: Drenkard, Eliana  
APPLICANT: Tsongalis, John  
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
FILE REFERENCE: 00786/361002  
CURRENT APPLICATION NUMBER: US/09/199,637A  
CURRENT FILING DATE: 1998-11-25  
PRIOR APPLICATION NUMBER: 60/066,517  
PRIOR FILING DATE: 1997-11-25  
NUMBER OF SEQ ID NOS: 437  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 111  
LENGTH: 6909  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-111

Query Match  
Best Local Similarity 82.2%; Score 14.8; DB 4; Length 6909;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGTGCGGGA 18  
|||||  
DB 2627 TGGCGCGGTGTGCGGGA 2644

RESULT 15  
US-08-611-757-14  
Sequence 14, Application US/08611757  
Patent No. 5859230  
GENERAL INFORMATION:  
APPLICANT: Kim, Jungsub P.  
APPLICANT: Reyes, Gregory R.  
APPLICANT: Wages, John  
APPLICANT: Zhang-Keck, Zhen-Yang  
APPLICANT: Young, Lavonne  
TITLE OF INVENTION: No. 5859230-A/No. 5859230-B/No. 5859230-C/No. 5859230-D/No. 58  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,757  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/246,985  
FILING DATE: 20-MAY-1994  
APPLICATION NUMBER: US 025,396  
FILING DATE: 24-FEB-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/922,493  
FILING DATE: 30-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 231 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: GB Clone 475-12-3  
US-08-611-757-14

Query Match  
Best Local Similarity 80.0%; Score 14.4; DB 2; Length 231;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCGCGGTGTGCGGGA 18  
|||||  
DB 131 GCGCGGTGTGCGGGA 146

Search completed: June 20, 2003, 23:25:44  
Job time: 14.1792 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:58:41 ; Search time 6.65094 Seconds  
(without alignments)  
3971.420 Million cell updates/sec

Title: US-09-965-553-2  
Perfect score: 18  
Sequence: 1 tggccgctgttcgggga 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCRUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	18	10	US-09-965-553-2
2	18	100.0	27	10	US-09-965-553-39
3	18	100.0	9829	10	US-09-965-553-19
4	18	100.0	10482	10	US-09-965-553-23
5	18	100.0	12286	10	US-09-965-553-17
6	16.4	91.1	2000	9	US-09-938-842A-5131
7	16.4	91.1	2364	9	US-09-712-363-32
8	15.4	85.6	251	10	US-09-878-574-10644
9	15.4	85.6	262	10	US-09-878-574-12308
10	15.4	85.6	269	10	US-09-878-574-12035
11	15.4	85.6	272	10	US-09-878-574-13422
12	15.4	85.6	273	10	US-09-878-574-11753
13	15.4	85.6	275	10	US-09-878-574-15624
14	15.4	85.6	275	10	US-09-878-574-12670
15	15.4	85.6	275	10	US-09-878-574-14365
16	15.4	85.6	276	10	US-09-878-574-8192
17	15.4	85.6	285	10	US-09-878-574-12806
18	15.4	85.6	449	9	US-09-918-995-28499
19	15.4	85.6	1649	9	US-09-997-816-1

20	14.8	82.2	128	10	US-09-864-761-26007	Sequence 26007, A
21	14.8	82.2	256	10	US-09-923-876-1817	Sequence 1817, Ap
22	14.8	82.2	496	10	US-09-918-995-401	Sequence 401, App
23	14.8	82.2	561	10	US-09-864-761-9555	Sequence 9555, Ap
24	14.8	82.2	944	10	US-09-974-300-5723	Sequence 5723, Ap
25	14.8	82.2	1018	10	US-09-452-239-35	Sequence 35, Appl
26	14.8	82.2	1078	9	US-09-452-239-41	Sequence 79, Appl
27	14.8	82.2	1110	9	US-10-121-988-79	Sequence 145, App
28	14.8	82.2	1110	9	US-10-121-988-145	Sequence 82, Appl
29	14.8	82.2	1113	9	US-10-121-988-82	Sequence 1258, Ap
30	14.8	82.2	1215	9	US-09-738-626-1258	Sequence 94, Appl
31	14.8	82.2	3348	9	US-10-138-838-94	Sequence 94, Appl
32	14.8	82.2	3348	9	US-10-139-031-94	Sequence 94, Appl
33	14.8	82.2	3348	9	US-10-138-905-94	Sequence 94, Appl
34	14.8	82.2	3348	9	US-10-138-916-94	Sequence 94, Appl
35	14.8	82.2	3348	9	US-09-976-800-94	Sequence 94, Appl
36	14.8	82.2	6909	9	US-09-975-719-111	Sequence 111, App
37	14.8	82.2	3309400	9	US-09-738-626-1	Sequence 1, Appl
38	14.4	80.0	189	10	US-09-821-167-8	Sequence 8, Appl
39	14.4	80.0	249	10	US-09-878-574-15547	Sequence 15547, A
40	14.4	80.0	264	10	US-09-878-574-15236	Sequence 15236, A
41	14.4	80.0	269	10	US-09-878-574-10867	Sequence 10867, A
42	14.4	80.0	271	10	US-09-878-574-14193	Sequence 14193, A
43	14.4	80.0	303	10	US-09-878-574-15351	Sequence 15351, A
44	14.4	80.0	334	10	US-09-878-574-688	Sequence 688, App
45	14.4	80.0	342	10	US-09-878-574-2125	Sequence 2125, Ap

#### ALIGNMENTS

RESULT 1  
US-09-965-553-2  
; Sequence 2, Application US/09965553  
; Patent No. US20020112259A1  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/965,553  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 09/322,478  
; PRIOR FILING DATE: 1999-05-28  
; PRIOR APPLICATION NUMBER: 60/087125  
; PRIOR FILING DATE: 1998-05-29  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 18  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-965-553-2

Query Match 100.0%; Score 18; DB 10; Length 18;  
Best Local Similarity 100.0%; Pred No. 5.2; Mismatches 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTCGGGGA 18  
|||||  
DB 1 TGGCGCCGTTGTCGGGGA 18

RESULT 2  
US-09-965-553-39  
; Sequence 39, Application US/09965553  
; Patent No. US20020112259A1  
; GENERAL INFORMATION:  
; APPLICANT: Wright, David A.  
; APPLICANT: Voytas, Daniel F.  
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
; FILE REFERENCE: P-1065 ISURF Plant Retroelement  
; CURRENT APPLICATION NUMBER: US/09/965,553

```

; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 39
; LENGTH: 27
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-39

Query Match      100.0%; Score 18; DB 10; Length 27;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
DB      5 TGGCGCCGTTGTGCGGGA 22

RESULT 3
US-09-965-553-19
; Sequence 19, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 19
; LENGTH: 9829
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-19

Query Match      100.0%; Score 18; DB 10; Length 9829;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
DB     1312 TGGCGCCGTTGTGCGGGA 1329

RESULT 4
US-09-965-553-23
; Sequence 23, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 23
; LENGTH: 10482
```

```

; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-23

Query Match      100.0%; Score 18; DB 10; Length 10482;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
DB     1448 TGGCGCCGTTGTGCGGGA 1465

RESULT 5
US-09-965-553-17
; Sequence 17, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 17
; LENGTH: 12286
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plant
US-09-965-553-17

Query Match      100.0%; Score 18; DB 10; Length 12286;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
DB     1312 TGGCGCCGTTGTGCGGGA 1329

RESULT 6
US-09-938-842A-5131
; Sequence 5131, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5131
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
```



US-09-938-842A-5131

Query Match 91.1%; Score 16.4; DB 9; Length 2000;  
Best Local Similarity 94.4%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGTCGGGGA 18  
|||||  
DB 1883 TGGCGCCGTTGTCGGGGA 1900

RESULT 7  
US-09-712-363-32/C

; Sequence 32, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 2364  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-32

Query Match 91.1%; Score 16.4; DB 9; Length 2364;  
Best Local Similarity 94.4%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGTCGGGGA 18  
|||||  
DB 1084 TGGCGCCGTTGTCGGGGA 1067

RESULT 8  
US-09-878-574-10644/C

; Sequence 10644, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535

; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 10644  
; LENGTH: 251  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 700967905H1  
US-09-878-574-10644

Query Match 85.6%; Score 15.4; DB 10; Length 251;  
Best Local Similarity 94.1%; Pred. No. 87;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GCGCGCCGTTGTCGGGGA 18  
|||||  
DB 243 GCGCGCCGTTGTCGGTGA 227

RESULT 9  
US-09-878-574-12308/C

; Sequence 12308, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 12308  
; LENGTH: 262  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 701065609H1  
US-09-878-574-12308

Query Match 85.6%; Score 15.4; DB 10; Length 262;  
Best Local Similarity 94.1%; Pred. No. 87;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GCGCGCCGTTGTCGGGGA 18  
|||||  
DB 244 GCGCGCCGTTGTCGGTGA 228

RESULT 10  
US-09-878-574-12035/C

; Sequence 12035, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 12035  
; LENGTH: 269  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 701065234H1  
US-09-878-574-12035

Query Match 85.6%; Score 15.4; DB 10; Length 269;  
Best Local Similarity 94.1%; Pred. No. 86;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18  
DB 243 GGCGCCGTTGTCGGTGA 227

## RESULT 11

US-09-878-574-13422/C  
Sequence 13422, Application US/09878574  
Patent No. US20020110548A1  
GENERAL INFORMATION:  
APPLICANT: Byrum, Joseph R.  
APPLICANT: La Rosa, Thomas J.  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(15401)B  
CURRENT APPLICATION NUMBER: US/09/878,574  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 09/333,535  
PRIOR FILING DATE: 1999-06-14  
NUMBER OF SEQ ID NOS: 15775  
SEQ ID NO 13422  
LENGTH: 269  
TYPE: DNA  
ORGANISM: Glycine max  
OTHER INFORMATION: Clone ID: 701067081H1  
US-09-878-574-13422

Query Match 85.6%; Score 15.4; DB 10; Length 269;  
Best Local Similarity 94.1%; Pred. No. 86;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18  
DB 268 GGCGCCGTTGTCGGTGA 252

## RESULT 12

US-09-878-574-11753/C  
Sequence 11753, Application US/09878574  
Patent No. US20020110548A1  
GENERAL INFORMATION:  
APPLICANT: Byrum, Joseph R.  
APPLICANT: La Rosa, Thomas J.  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(15401)B  
CURRENT APPLICATION NUMBER: US/09/878,574  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 09/333,535  
PRIOR FILING DATE: 1999-06-14  
NUMBER OF SEQ ID NOS: 15775  
SEQ ID NO 11753  
LENGTH: 272  
TYPE: DNA  
ORGANISM: Glycine max  
OTHER INFORMATION: Clone ID: 701064849H1  
US-09-878-574-11753

Query Match 85.6%; Score 15.4; DB 10; Length 272;  
Best Local Similarity 94.1%; Pred. No. 86;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18  
DB 226 GGCGCCGTTGTCGGTGA 210

RESULT 13  
US-09-878-574-15624/C  
Sequence 15624, Application US/09878574  
Patent No. US20020110548A1  
GENERAL INFORMATION:  
APPLICANT: Byrum, Joseph R.  
APPLICANT: La Rosa, Thomas J.  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(15401)B  
CURRENT APPLICATION NUMBER: US/09/878,574  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 09/333,535  
PRIOR FILING DATE: 1999-06-14  
NUMBER OF SEQ ID NOS: 15775  
SEQ ID NO 15624  
LENGTH: 273  
TYPE: DNA  
ORGANISM: Glycine max  
OTHER INFORMATION: Clone ID: 701070324H1  
US-09-878-574-15624

Query Match 85.6%; Score 15.4; DB 10; Length 273;  
Best Local Similarity 94.1%; Pred. No. 86;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18  
DB 239 GGCGCCGTTGTCGGTGA 223

## RESULT 14

US-09-878-574-12670/C  
Sequence 12670, Application US/09878574  
Patent No. US20020110548A1  
GENERAL INFORMATION:  
APPLICANT: Byrum, Joseph R.  
APPLICANT: La Rosa, Thomas J.  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(15401)B  
CURRENT APPLICATION NUMBER: US/09/878,574  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 09/333,535  
PRIOR FILING DATE: 1999-06-14  
NUMBER OF SEQ ID NOS: 15775  
SEQ ID NO 12670  
LENGTH: 275  
TYPE: DNA  
ORGANISM: Glycine max  
OTHER INFORMATION: Clone ID: 701066064H1  
US-09-878-574-12670

Query Match 85.6%; Score 15.4; DB 10; Length 275;  
Best Local Similarity 94.1%; Pred. No. 86;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18  
DB 235 GGCGCCGTTGTCGGTGA 219

## RESULT 15

US-09-878-574-14365/C  
Sequence 14365, Application US/09878574  
Patent No. US20020110548A1  
GENERAL INFORMATION:  
APPLICANT: Byrum, Joseph R.  
APPLICANT: La Rosa, Thomas J.  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

```
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 14365
; LENGTH: 275
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701068668H1
; US-09-878-574-14365

Query Match      85.6%; Score 15.4; DB 10; Length 275;
Best Local Similarity 94.1%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      2 GGCGCCGTTGTCGGGGA 18
        |||||||||||||
Db      246 GGCGCCGTTGTCGGTGA 230
```

Search completed: June 21, 2003, 00:34:40  
Job time : 8.65094 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:49:21 ; Search time 69.1132 Seconds  
(without alignments)  
4217.989 Million cell updates/sec

Title: US-09-965-553-2  
Perfect score: 18  
Sequence: 1 tggcgcgtgttcgggga 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estha:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_huv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_trod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	250	17	AQ288973 nbxb0033H
2	18	100.0	552	17	BH731480 BOHZ016TF
3	18	100.0	624	17	BH545031 BOGCW07TF
4	18	100.0	668	17	BH600565 BOGDC03TR
5	18	100.0	678	17	BH214458 Gm_DMD001
6	18	100.0	690	17	BH484542 BOGWE45TF

7	18	100.0	694	17	BH725980 BOMMT13TR
8	18	100.0	697	13	BT642756 T2S973 T2
9	18	100.0	708	17	BH552125 BOHES25TF
10	18	100.0	720	17	BH445346 BOHNP83TR
11	18	100.0	722	17	BH424623 BOGEE39TR
12	18	100.0	741	17	BH141140 Gm_DMD001
13	18	100.0	747	17	BH505910 BOGDT62TR
14	18	100.0	768	17	BH590633 BOGHB52TF
15	18	100.0	770	17	BH423648 BOGMY06TR
16	18	100.0	771	17	BH700732 BOMN242TF
17	18	100.0	782	12	BG299448 HVSME002
18	18	100.0	789	17	BH724831 BOMLV90TR
19	18	100.0	796	17	BH514436 BOGUV17TR
20	18	100.0	805	17	BH531932 BOGMR94TF
21	18	100.0	809	17	BH420835 BOGCM07TR
22	18	100.0	819	17	BH454350 BOGBJ86TR
23	18	100.0	825	12	BF620501 HVSMEC002
24	18	100.0	828	17	BH583501 BOGHN35TR
25	18	100.0	848	17	BH451469 BOGXH39TF
26	18	100.0	858	17	BH698149 BOMNR68TF
27	18	100.0	864	17	BH246588 BOGAP59TR
28	17	94.4	719	17	BH591785 BOHAJ29TR
29	16.4	91.1	85	17	A2922024 HRCOC2B09
30	16.4	91.1	147	10	A2921883 HRCOC3E09
31	16.4	91.1	159	17	BE086799 OY1-BT067
32	16.4	91.1	159	17	AM679402 WSL_24_CO
33	16.4	91.1	159	17	BH650606 BOGMR296TF
34	16.4	91.1	160	17	A2922212 MKCOC2H07
35	16.4	91.1	160	17	BH501549 BOHMM57TR
36	16.4	91.1	231	17	A2922191 MKCOC2D08
37	16.4	91.1	238	17	AL094958 Arabidops
38	16.4	91.1	282	17	BH436253 BOGCH70TF
39	16.4	91.1	288	17	BH744134 g3c2c02.D
40	16.4	91.1	291	17	BH705388 BOHVA74TR
41	16.4	91.1	300	17	BH745888 g273608.D
42	16.4	91.1	323	17	BH514295 BOGJO18TR
43	16.4	91.1	326	17	BH557816 BOGXT09TF
44	16.4	91.1	334	17	BH700667 BOMHO22TF
45	16.4	91.1	336	17	BH727785 BOHZN26TF

## ALIGNMENTS

RESULT 1  
LOCUS AQ288973 250 bp DNA linear GSS 03-DEC-1998  
DEFINITION nbxb0033H14f CUGI Rice BAC Library Oryza sativa genomic clone  
ACCESSION AQ288973.1 GI:3950419  
VERSION AQ288973.1  
KEYWORDS GSS.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.

REFERENCE  
AUTHORS Wing, R.A. and Dean, R.A.  
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome  
JOURNAL Unpublished (1998)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATGACCTCAGCTATAGGG  
Class: BAC ends  
High quality sequence stop: 117.  
Location/Qualifiers  
1. 250

FEATURES  
source

```

/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone_1lb="CUG1 Rice BAC library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBelobAC11; Site.1: HindIII; Site.2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT      88 a      49 c      59 g      54 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 250;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCCGTTGCGGGA 18
        ||||||||||||||||
Db      165 TGGCGCCGTTGCGGGA 148

RESULT 2
LOCUS      BH731480      552 bp      DNA      linear      GSS 20-FEB-2002
DEFINITION BOH2016F BO_2_3_KB Brassica oleracea genomic clone BOH2016, DNA
sequence.
ACCESSION      BH731480
VERSION      BH731480.1 GI:18836875
KEYWORDS      GSS.
SOURCE      Brassica oleracea.
ORGANISM      Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 552)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOH2016TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES
source
1..552
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_1lb="BO_2_3_KB"
/clone_1lb="BO_2_3_KB"

```

```

/notes="Vector: PHOS1; Site.1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT      154 a      111 c      97 g      190 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 552;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCCGTTGCGGGA 18
        ||||||||||||||||
Db      364 TGGCGCCGTTGCGGGA 381

RESULT 3
LOCUS      BH545031      624 bp      DNA      linear      GSS 14-DEC-2001
DEFINITION BOGCW07TF BOGC Brassica oleracea genomic clone BOGCW07, DNA
sequence.
ACCESSION      BH545031
VERSION      BH545031.1 GI:17796812
KEYWORDS      GSS.
SOURCE      Brassica oleracea.
ORGANISM      Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 624)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOGCW07TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES
source
1..624
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_1lb="BOGCW07"
/clone_1lb="BOGC"
/notes="Vector: PHOS1; Site.1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT      185 a      129 c      109 g      201 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 624;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCCGTTGCGGGA 18
        ||||||||||||||||
Db      542 TGGCGCCGTTGCGGGA 559

RESULT 4
LOCUS      BH600565/c      668 bp      DNA      linear      GSS 15-DEC-2001
DEFINITION BOGDG03TR BOGD Brassica oleracea genomic clone BOGDG03, DNA
sequence.
ACCESSION      BH600565
VERSION      BH600565.1 GI:17853011
KEYWORDS      GSS.
SOURCE      Brassica oleracea.
ORGANISM      Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

REFERENCE  
 TITLES  
 JOURNAL  
 COMMENT

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 668)  
 Town, C.D., Van Aken, S., Utterback, T., and Fraser, C.M. Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)  
 Other GSSs: BOGD037F  
 Contact: Chris Town

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 Location/Qualifiers

1..668  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOGD03"

BASE COUNT  
 ORIGIN

186 a 146 c 144 g 192 t

Query Match  
 Best Local Similarity 100.0%; Score 18; DB 17; Length 668;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCGCCGCTGTGCGGGA 18  
 Db 523 TGCGCCGCTGTGCGGGA 506

RESULT 5  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BH214458 678 bp DNA linear GSS 25-OCT-2001  
 Gm\_UMB001\_098\_A01.F UMN Soybean BAC Library (PECSBAC4 EcorI)  
 Glycine max genomic clone glycine max genomic clone  
 Gm\_UMB001\_098\_A01, DNA sequence.  
 BH214458  
 BH214458.1 GI:16428384  
 GSS.  
 soybean.  
 Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine. 1 (bases 1 to 678)  
 1 (bases 1 to 678)  
 Marek, L.F., Paz, M., Darnelle, L., Hanson, N. and Shoemaker, R.C. BAC End sequences from a soybean genomic library (ISU) unpublished (2000)  
 Contact: Shoemaker Randy C  
 Agronomy Department  
 Iowa State University  
 Ames, IA 50011-1010, USA  
 Tel: 515 294 1205  
 Fax: 515 294 2299  
 Email: rcsshoe@iastate.edu  
 This BAC identified by SSR Satt077. For more information, see Soybase at:  
 http://genome.cornell.edu/cgi-bin/webace/webace?db=soybase. Please see as an authority for the mapping/naming: Cregan P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Iark, A.L. Kahler, N. Kaya, T.T. Vantoui, D.G. Lohnes, J. Chung, and J.E. Specht. 1999a. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464-1490.  
 Seg primer: M13F  
 Class: BAC ends.

FEATURES  
 source

Location/Qualifiers  
 1..678  
 /organism="Glycine max"  
 /cultivar="Faribault"  
 /db\_xref="taxon:3847"  
 /clone="Gm\_UMB001\_098\_A01"  
 /clone\_lib="UMN Soybean BAC Library (PECSBAC4 EcorI)  
 Glycine max genomic clone"  
 /tissue\_type="cotyledon leaves"  
 /dev\_stage="cotyledon"  
 /note="Vector: pECSBAC4; the UMN BAC library (Danesh et al Theor. Appl. Genet. 96:196, 1998) was constructed using the Eco RI site of pECSBAC4. The library consists of 72,960 clones with an average insert size of 120 Kb, equal to 7 haploid genome equivalents. Screening of the library is done by hybridization of high-density colony filters and/or PCR amplification of DNA pools. Four high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening"

BASE COUNT  
 ORIGIN

180 a 92 c 80 g 301 t 25 others

Query Match  
 Best Local Similarity 100.0%; Score 18; DB 17; Length 678;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCGCCGCTGTGCGGGA 18  
 Db 539 TGCGCCGCTGTGCGGGA 556

RESULT 6  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BH484542 690 bp DNA linear GSS 13-DEC-2001  
 BOGWE45TF BOGW Brassica oleracea genomic clone BOGWE45, DNA sequence.  
 BH484542  
 BH484542.1 GI:17692646  
 GSS.  
 Brassica oleracea.  
 Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 690)  
 1 (bases 1 to 690)  
 Town, C.D., Van Aken, S., Utterback, T., and Fraser, C.M. Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)  
 Other GSSs: BOGWE45TF  
 Contact: Chris Town

REFERENCE  
 TITLES  
 JOURNAL  
 COMMENT

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TF  
 Class: sheared ends.

FEATURES  
 source  
 Location/Qualifiers

1..690  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOGWE45"  
 /clone\_lib="BOGW"  
 /note="Vector: PHOS1, site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT  
 ORIGIN

221 a 139 c 140 g 180 t

Query Match  
 Best Local Similarity 100.0%; Score 18; DB 17; Length 690;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCGGGA 18  
 |||  
 Db 280 TGGCGCCGTTGCGGGA 263

RESULT 7  
 BH725980 694 bp DNA linear GSS 20-FEB-2002  
 LOCUS BOMWT13R BO\_2\_3-KB Brassica oleracea genomic clone BOMWT13, DNA  
 DEFINITION  
 accession  
 sequence.  
 ACCESSION BH725980  
 VERSION BH725980.1 GI:18631375  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea.  
 ORGANISM Brassica oleracea.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 694)  
 Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 Unpublished (2001)  
 Other-GSSs: BOMWT13TF  
 COMMENT  
 Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 1..694  
 Location/Qualifiers  
 /organism="Brassica oleracea"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOMWT13"  
 /clone\_1lb="BO\_2\_3-KB"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared  
 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 188 a 182 c 139 g 185 t

ORIGIN  
 Query Match 100.0%; Score 18; DB 17; Length 694;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCGGGA 18  
 |||  
 Db 133 TGGCGCCGTTGCGGGA 150

RESULT 8  
 BI642756 697 bp mRNA linear EST 10-SEP-2001  
 LOCUS T2S973 T2S (Sapwood-heartwood transition zone of black locust -  
 DEFINITION Summer) Robinia pseudoacacia cDNA, mRNA sequence.  
 accession  
 sequence.  
 ACCESSION BI642756  
 VERSION BI642756.1 GI:115544966  
 KEYWORDS EST.  
 SOURCE Robinia pseudoacacia.  
 ORGANISM Robinia pseudoacacia.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Robinieae;  
 Robinia.  
 1 (bases 1 to 697)  
 Han,K.-H., Yang,J., Park,S., Paule,C.R., Kapur,V., Retzel,E.F.,  
 Kamdem,D.P. and Keathley,D.E.  
 Analysis of gene expression patterns in trunk wood of a mature  
 black locust (Robinia pseudoacacia)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

JOURNAL  
 COMMENT Unpublished (2002)  
 Contact: Kyung-Hwan Han  
 Department of Forestry  
 Michigan State University  
 126 Natural Resources, East Lansing, MI 48824-1222, USA  
 Tel: 517 353 4751  
 Fax: 517 432 1143  
 Email: hanky@msu.edu.

FEATURES  
 source  
 1..697  
 Location/Qualifiers  
 /organism="Robinia pseudoacacia"  
 /db\_xref="taxon:35938"  
 /clone\_1lb="T2S (Sapwood-heartwood transition zone of  
 black locust - Summer)"  
 /tissue\_type="sapwood-heartwood transition zone"  
 /dev\_stage="mature tree"  
 /note="Vector: lambda Triplex; Site\_1: Sfi IA; Site\_2: Sfi  
 IB; The cDNA library was made from the sapwood-heartwood  
 transition zone of a mature black locust tree collected in  
 Michigan in late July." 123 c 94 g 208 t

BASE COUNT 272 a 123 c 94 g 208 t

ORIGIN  
 Query Match 100.0%; Score 18; DB 13; Length 697;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCGGGA 18  
 |||  
 Db 25 TGGCGCCGTTGCGGGA 8

RESULT 9  
 BH552125 708 bp DNA linear GSS 14-DEC-2001  
 LOCUS BOHES25TF BOHE Brassica oleracea genomic clone BOHES25, DNA  
 DEFINITION  
 sequence.  
 accession  
 sequence.  
 ACCESSION BH552125  
 VERSION BH552125.1 GI:17803905  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea.  
 ORGANISM Brassica oleracea.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 708)  
 Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 Unpublished (2001)  
 Other-GSSs: BOHES25TR  
 COMMENT  
 Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TF  
 Class: sheared ends.

FEATURES  
 source  
 1..708  
 Location/Qualifiers  
 /organism="Brassica oleracea"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOHES25"  
 /clone\_1lb="BOHE"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared  
 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 191 a 181 c 138 g 198 t

ORIGIN  
 Query Match 100.0%; Score 18; DB 17; Length 708;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;



Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGCGCCGTGTGCGGGA 18  
 ||||||||||||||||  
 Db 63 TGGCGCCGTGTGCGGGA 80

RESULT 10  
 BH445346 720 bp DNA linear GSS 12-DEC-2001  
 LOCUS BOHNP83TR BOHN Brassica oleracea genomic clone BOHNP83, DNA  
 DEFINITION

ACCESSION BH445346  
 VERSION BH445346  
 KEYWORDS GI:17631060  
 SOURCE GSS.  
 ORGANISM Brassica oleracea.

REFERENCE  
 AUTHORS Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.  
 TITLE Whole genome shotgun sequencing of Brassica oleracea  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: BOHNP83TF  
 CONTACT: Chris Town

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 Location/Qualifiers

1..720  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOHNP83"  
 /note="Vector: PHOS1, Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 204 a 179 c 135 g 202 t  
 ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 720;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGCGCCGTGTGCGGGA 18  
 ||||||||||||||||  
 Db 221 TGGCGCCGTGTGCGGGA 238

RESULT 11  
 BH424623 722 bp DNA linear GSS 12-DEC-2001  
 LOCUS BOEEB39TR BOEE Brassica oleracea genomic clone BOEEB39, DNA  
 DEFINITION

ACCESSION BH424623  
 VERSION BH424623  
 KEYWORDS GI:17610351  
 SOURCE GSS.  
 ORGANISM Brassica oleracea.

REFERENCE  
 AUTHORS Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.  
 TITLE Whole genome shotgun sequencing of Brassica oleracea  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: BOEEB39TF

Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 Location/Qualifiers

1..722  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOEEB39"  
 /note="Vector: PHOS1, Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 254 a 132 c 110 g 226 t  
 ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 722;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGCGCCGTGTGCGGGA 18  
 ||||||||||||||||  
 Db 684 TGGCGCCGTGTGCGGGA 701

RESULT 12  
 BH141140 741 bp DNA linear GSS 15-AUG-2001  
 LOCUS Gm\_UMD001\_098\_A01F UMN Soybean BAC Library (PECSBAC4 EcORI) Glycine  
 DEFINITION max genomic clone Glycine max genomic, DNA sequence.

ACCESSION BH141140  
 VERSION BH141140.1 GI:15187351  
 KEYWORDS GSS.  
 SOURCE soybean.  
 ORGANISM Glycine max

REFERENCE  
 AUTHORS Larsen, D., Mudge, D., Denny, R., Yan, H., Danesh, D. and Young, N.D.  
 TITLE BAC end sequences from a soybean BAC genomic library (UMN)  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Young Nevlin D

Department of Plant Pathology  
 University of Minnesota  
 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul MN 55108, USA  
 Tel: 612 625 2225  
 Fax: 612 625 9728  
 Email: nevin@tc.umn.edu  
 Sequence on contig Gm\_A455.ctg a near unmapped duplicate of pA455.  
 For more information, see SoyBase at:  
<http://genome.cornell.edu/cgi-bin/WebBase/webacer7db-soybaseclass-locus>.  
 Please see as an authority for the mapping/naming: Cregan P.B.,  
 T. Jarvik, A.L. ush, R.C. Shoemaker, K.G. Iark, A.L. Kahler, N.  
 Kaya, T.T. Vantolai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999a.  
 An integrated genetic linkage map of the soybean genome. Crop Sci.  
 39:1464-1490  
 Seq primer: M13P  
 Class: BAC ends.

FEATURES  
 Location/Qualifiers

1..741  
 /organism="Glycine max"  
 /cultivar="Faribault"  
 /db\_xref="taxon:3847"  
 /clone\_lib="UMN Soybean BAC library (PECSBAC4 EcORI)"  
 Glycine max genomic clone"  
 /tissue\_type="cotyledon leaves"

```

/dev-stage="cotyledon"
/Note="Vector: pECSBAC4; The UMN BAC library (Danesh et al
Theor. Appl. Genet. 96:196, 1998) was constructed using
the Eco RI site of pECSBAC4. The library consists of 72
,960 clones with an average insert size of 120 kb, equal
to 7 haploid genome equivalents. Screening of the library
is done by hybridization of high-density colony filters
and/or PCR amplification of DNA pools. Four high density
filters, each containing 18,432 clones (doubly spotted),
represent the whole library for colony screening"

BASE COUNT      200 a      97 c      82 g      333 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 741;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
Db      548 TGGCGCCGTTGTGCGGGA 565

RESULT 13
BH505910      747 bp      DNA      linear      GSS 13-DEC-2001
DEFINITION    BCGDT64TR BCGD Brassica oleracea genomic clone BCGDT64, DNA
sequence.
ACCESSION     BH505910
VERSION       BH505910.1 GI:17714007
KEYWORDS      GSS.
SOURCE        Brassica oleracea.
ORGANISM      Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 747)
TOWN,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
unpublished (2001)
Other-GSSs: BCGDT64TF
Contact: Chris Town
TIGR          9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1. 747
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGDT64"
/clone_lib="BOGD"
/Note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT      214 a      188 c      136 g      209 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 747;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
Db      233 TGGCGCCGTTGTGCGGGA 250

RESULT 14
BH590633/c      768 bp      DNA      linear      GSS 15-DEC-2001
LOCUS

```

```

DEFINITION      BOHE252TR BOHB Brassica oleracea genomic clone BOHB252, DNA
sequence.
ACCESSION       BH590633
VERSION         BH590633.1 GI:17843085
KEYWORDS        GSS.
SOURCE          Brassica oleracea.
ORGANISM        Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 768)
TOWN,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
unpublished (2001)
Other-GSSs: BOHB252TR
Contact: Chris Town
TIGR          9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1. 768
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHB252"
/clone_lib="BOHB"
/Note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT      207 a      169 c      179 g      213 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 768;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
Db      651 TGGCGCCGTTGTGCGGGA 634

RESULT 15
BH423648/c      770 bp      DNA      linear      GSS 12-DEC-2001
LOCUS          BCGWY06TR BCGW Brassica oleracea genomic clone BCGWY06, DNA
sequence.
ACCESSION       BH423648
VERSION         BH423648.1 GI:17609376
KEYWORDS        GSS.
SOURCE          Brassica oleracea.
ORGANISM        Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 770)
TOWN,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
unpublished (2001)
Other-GSSs: BCGWY06TF
Contact: Chris Town
TIGR          9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers

```

```
source          1. 770
                 /organism="Brassica oleracea"
                 /strain="T0100DH3"
                 /db_xref="taxon:3712"
                 /clone="BOGWY06"
                 /clone_1lb="BOGW"
                 /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
BASE COUNT      220 a 130 c 167 g 253 t
ORIGIN
Query Match     100.0%; Score 18; DB 17; Length 770;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TGGCGCGGTGTGCGGGA 18
    ||||||||||||||||
Db 526 TGGCGCGGTGTGCGGGA 509
```

Search completed: June 20, 2003, 22:39:16  
Job time : 73.1132 secs

**THIS PAGE BLANK (USPTO)**